

SEARCH REQUEST FORM
Scientific and Technical Information Center

Access DB#

53593

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr. Link _____
Date Completed: <u>10/30/01</u>	Litigation _____	Lexis Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Client Prep Time: _____	Patent Family _____	WWW Internet _____
Online Time: _____	Other _____	Other (specify): _____

Contact:
Sheppard

tel: 308-4499

~~STIC~~-Biotech/ChemLib

53593

From: Prouty, Rebecca
Sent: Wednesday, October 24, 2001 8:44 AM
To: STIC-Biotech/ChemLib
Subject: Sequence Search

Art Unit 1652, 10A13
Mailbox: 10C01
308-4000
Serial Number: 09/416,579

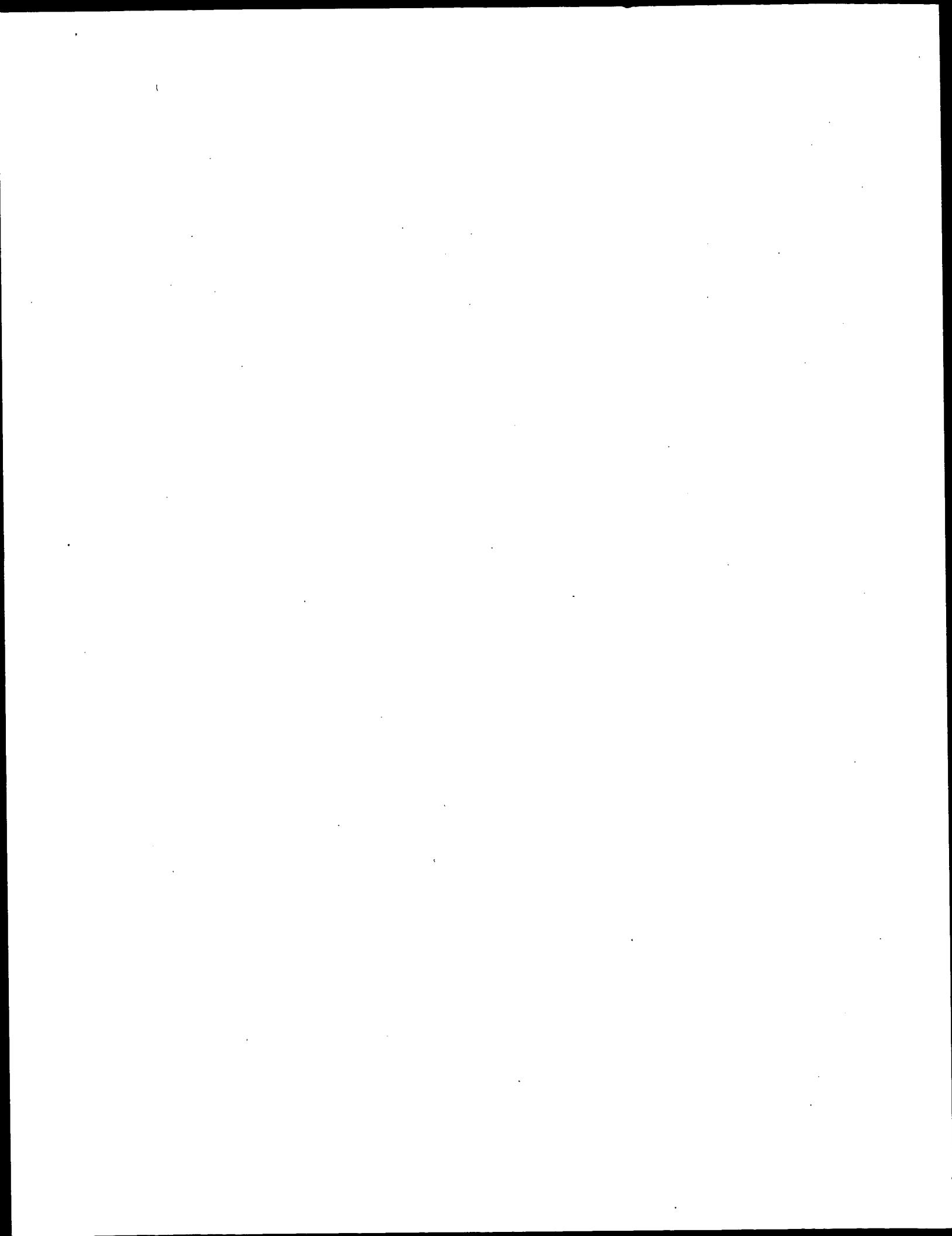
Please search and interference search SEQ ID NO:1 and the translation of SEQ NO:1

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 03:57:24 ; Search time 1156.6 Seconds
(without alignments)
6154.244 Million cell updates/sec

Title: US-09-416-579A-1

Perfect score: 753

Sequence: 1 atggcggagggagcagctctg.....gccagagggtccagagtaa 753

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 segs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing First 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	465.2	61.8	565	21	AT514891 LD41330.5
2	464.8	61.7	597	21	AT512995 LD45202.5
3	453.2	60.2	563	10	AA697070 GM09038.5
4	452.8	59.3	628	14	AA979445 LD34014.5
5	446.2	59.3	590	7	AA441228 LD15983.5
6	412	54.7	653	14	AA949518 LD28349.5
7	186.8	24.8	712	106	AU004911 AU004911
8	144.4	19.2	689	30	AV404522 AV404522
9	114.8	15.2	986	106	AL583655 AL583655
10	94.8	12.6	899	146	BF233387 BF233387
11	85.2	11.3	763	107	AU050828 AU050828
12	83	11.0	646	112	AM184455 f114d08.Y
13	82.2	10.9	294	151	BF652168 275654.MA
14	77.8	10.3	531	113	AM226818 um62f04.Y
15	77.8	10.3	664	163	BE132294 BE132294
16	74.6	9.9	334	13	AA920068 vY67h10.r
17	74.6	9.9	782	107	AU035388 AU035388
18	70.4	9.3	601	111	AM106746 AM106746
19	66	8.8	513	22	AT629362 fC05e10.Y
20	65.6	8.7	455	153	BG408500 d402a05.Y
21	63.8	8.5	419	106	AL586236 AL586236
22	63.2	8.4	1025	174	BE176406 BE176406
23	62.4	8.4	1967	144	BF120228 BF120228
24	62.2	8.3	381	114	AW323769 u076d01.Y
25	57.8	7.7	493	143	BF024786 d411a04.Y
26	54.6	7.3	260	168	BF712055 MI-P-E6-a
27	54	7.2	464	112	AM159435 z49b12.x
28	51.6	6.9	760	111	AM058933 f450h07.x
29	49.2	6.5	524	174	BE145038 u088a08.Y
30	48.4	6.4	355	174	BE145572 mac29e02.
31	48.4	6.4	440	168	BF715304 m406g11.
32	48.4	6.4	438	138	BE623416 u477a07.Y
33	48.4	6.4	526	149	BF450627 u278d11.Y
34	48.4	6.4	627	121	AM822817 uq20b01.Y
35	48.4	6.4	680	143	BF023209 uX05e09.Y
36	48.4	6.4	1003	192	AK010700 Mus muscu
37	48.4	6.4	2880	192	AK004592 Mus muscu
38	48	6.4	702	173	BG095503 u088a08.x
39	47.4	6.3	414	4	AA285789 vB87f04.r
40	46.8	6.2	538	31	AV628158 AV628158
41	46.8	6.2	604	120	AM741020 uq28h03.Y
42	46.4	6.2	849	220	CNS02JLA AL202951 Tetradon
43	46.2	6.1	587	23	AL657604 f415f06.Y
44	46.2	6.1	667	148	BF404976 UI-R-CAL-
45	46.2	6.1	716	104	AB392458 AB392458

ALIGNMENTS

RESULT 1

LOCUS AT514891 565 bp mRNA EST 16-MAR-1999

DEFINITION LD41330.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD41330 5prime, mRNA sequence.

ACCESSION AT514891

VERSION AT514891.1 GI:4418789

KEYWORDS EST.

SOURCE fruit fly, *Drosophila melanogaster*

ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 565)

AUTHORS Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Broksstein 'P., Lewis,S. and Rubin,G.M.

TITLE BDBP/HHMI Drosophila EST project

JOURNAL COMMENT

Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: 413 row: C column: 6
High quality sequence stop: 458.
Location/Qualifiers

FEATURES

source

1. 565

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="LD41330"

/clone_lib="LD Drosophila melanogaster embryo pot2"

/sex="male and female"

/dev_stage="0 to 24 hours mixed stage embryonic"

/lab_host="X1 Blue"

/note="Organ: embryo; Vector: pot2; Site:1; EcoRI; Site:2; XhoI: Sized fractionated cDNAs were directly ligated into pot2."

BASE COUNT

149 a 159 c 146 g 111 t

ORIGIN

Query Match

Best Local Similarity 99.4%; DB 21; Length 565;

Matches 467; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atggcggagcagcactcctctgcccgaaggaggaacagcagcagcagcagc 60

DB 96 atggcggagcagcactcctctgcccgaaggaggaacagcagcagcagcagc 155

QY 61 ttacacgtcctcctcagagggcaactcggcggaggaagcagcagcagcagc 120

DB 156 ttacacgtcctcctcagagggcaactcggcggaggaagcagcagcagcagc 215

QY 121 ggaagatcaagaacagacattcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 180

DB 216 ggaagatcaagaacagacattcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 275

QY 181 aacgggtaactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 240

DB 276 aacgggtaactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 335

QY 241 cagagttatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 300

DB 336 cagagttatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 395

QY 301 aaataargagcgtcctccttcttctcgtcgtcgtcgtcgtcgtcgtcgtcgtc 360

DB 396 aaataargagcgtcctccttcttctcgtcgtcgtcgtcgtcgtcgtcgtcgtc 455

QY 361 aacgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 420

DB 456 aacgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 515

QY 421 gagtcattcattcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 470

DB 516 gagtcattcattcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 565

RESULT 2

LOCUS AT512995 597 bp mRNA EST 16-MAR-1999

DEFINITION LD45202.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD45202 3prime, mRNA sequence.

ACCESSION AT512995

VERSION AT512995.1 GI:4422413

KEYWORDS EST.

SOURCE fruit fly, *Drosophila melanogaster*

ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;


```

RESULT 4
AA979445 628 bp mRNA EST 24-NOV-1998
LOCUS LD34014.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
DEFINITION melanogaster cDNA clone LD34014 5prime, mRNA sequence.
ACCESSION AA979445
VERSION AA979445.1 GI:3156850
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE Drosophila melanogaster
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 628)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
CONTACT: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 340 row: B column: 2
High quality sequence stop: 616.
FEATURES
source
1..628
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD34014"
/clone_1lb="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."
BASE COUNT 163 a 178 c 164 g 123 t
ORIGIN
Query Match 60.1%; Score 452.8; DB 14; Length 628;
Best Local Similarity 99.2%; Pred. No. 3.5e-112;
Matches 476; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 1 atggcgaggagcagcatcctgtgcccgaagggagaccagtaagcgagggcaccgcc 60
|||||
DB 151 atggcgaggagcagcatcctgtgcccgaagggagaccagtaagcgagggcaccgcc 209
QY 61 ttaccgctccatcgaggagcaacatcgagcgaggaagaccagtaattgaacccttc 120
|||||
DB 210 ttaccgctccatcgaggagcaacatcgagcgaggaagaccagtaattgaacccttc 269
QY 121 gagaagtaacaagaacacattgcctgctgacgcagccgcgtcgagaagtgcgcaagtc 180
|||||
DB 270 gagaagtaacaagaacacattgcctgctgacgcagccgcgtcgagaagtgcgcaagtc 329
QY 181 aacggggttaaatctgtgagctgagctgatacaagaatcccaagaagtgtgcccattc 240
|||||
DB 330 AACGGGTTAAATCTGCTGAGCTGATGTACAAAGATCCCAAGAGTGGCCATGCCCTTT 389
QY 241 cagaagtatgtacagctgacccaatgctgacgacacaccccccacaacaagaagctc 300
|||||
DB 390 CAGAGATT-TGTACAGCTGTACATGTGCTGACACACCGCCCAACACAAAGAGCTA 448
QY 301 aaataaargaacgctccttctttagcgtcgtatcttctgtaggaagaatgcgacga 360
|||||
DB 449 AAATATATGACGCTCCATTTTATGCGCTCGCTATTGCTTGTGAGAGACATGCGACGA 508
QY 361 aacggtcgtgtaggagcgatgtacaatacagctcgaggaagtgtgtacaagtctatcgaa 420

```

```

DB 509 AACGGCTCCCTGGAGCAGGACATGTACATACCTGGAGAGTGGTACAAAGTTCAATCAA 568
|||||
QY 421 gattccattaccctgcaagcgaccatcatatctcgcgaccctgcggaagtcgacg 480
|||||
DB 569 GAGTCATTCACGTGAGCGAGCGACCTCATCATATATCTGCGCACTCCCGAGAGTGGCG 628
|||||
RESULT 5
AA441228 590 bp mRNA EST 27-NOV-1998
LOCUS LD15983.5prime LD Drosophila melanogaster embryo Bluescript
DEFINITION Drosophila melanogaster cDNA clone LD15983 5prime, mRNA sequence.
ACCESSION AA441228
VERSION AA441228.1 GI:2794077
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE Drosophila melanogaster
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 590)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
CONTACT: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 159 row: G column: 11
High quality sequence stop: 412.
FEATURES
source
1..590
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="BDGP-EST:BDc1n015118"
/db_xref="taxon:7227"
/clone="LD15983"
/clone_1lb="LD Drosophila melanogaster embryo Bluescript"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI
; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit; Oligo dT-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT 158 a 165 c 150 g 117 t
ORIGIN
Query Match 59.3%; Score 446.2; DB 7; Length 590;
Best Local Similarity 99.1%; Pred. No. 2.1e-110;
Matches 459; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 atggcgaggagcagcatcctgtgcccgaagggagaccagtaagcgagggcaccgcc 60
|||||
DB 128 atggcgaggagcagcatcctgtgcccgaagggagaccagtaagcgagggcaccgcc 187
QY 61 ttaccgctccatcgaggagcaacatcgagcgaggaagaccagtaattgaacccttc 120
|||||
DB 188 TTCAACCGCTCATGAGGAGGCAACATGCGCAGCGGAAAGACACAGTATTGTAACCACTTC 247
QY 121 gagaagtaacaagaacacattgcctgctgacgcagccgcgtcgagaagtgcgcaagtc 180
|||||
DB 248 GAGAAGTACAAAGACGACATTTGCTGCTGACCGAGCCGTCGAGAAGTGGCCCAAGCTC 307
QY 181 aacggggttaaatctgtgagctgagctgatacaagaatcccaagaagtgtgcccattc 240
|||||
DB 308 AACGGGTTAAATCTGCTGAGCTGATGTACAAAGATCCCAAGAGTGGCGACCTCTTT 367

```

QY	241	caaggtatgtccagctgacacgtacacgtcgtgagtcgacacacccgccaacaaagaagcta	300
Db	368	CAGAGTATATGACGCTGACGTCACATGCTGACATGCGACACCCGCCAACCAAGAAGCTA	427
QY	301	aaataaagagagcgc-tccatttttaagcgcctgcgtcattgttcctgtgagaacatgacgc	359
Db	428	AAAATGAATGAGCGCGTTCATTTTAAAGCGCTCGCATATGCTTCGTGGAACATGCGACG	487
QY	360	aaagcgtcgtctgtagacagcaggcagtgataacgtctgtagaagatggttaacagttcatga	419
Db	488	AAAGCGCTCGCTGGAGCGACGGGCAATGTAAATAAGCTGGAGGAGTGTCAAGATTTCATCGA	547
QY	420	agaagtcattcacctgcagcgcggaacctcatatataatctgtgc 462	
Db	548	AGAGTCATTCACGTCGACGAGCGGACCTCATCATATATCTGCGC 590	
RESULT	6		
LOCUS	AA949518	653 bp	mRNA
DEFINITION	LD28349.5prlme LD Drosophila melanogaster embryo pot2 Drosophila		EST
ACCESSION	AA949518		25-NOV-1998
VERSION	AA949518.1	GI:3111350	
KEYWORDS	EST.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
	1 (bases 1 to 653)		
	Harvey, D., Hong, J., Evans-Holm, M., Pendleton, J., Su, C., Brokstein		
	P., Lewis, S., and Rubin, G. M.		
	BDGP/HMI Drosophila EST Project		
	Unpublished (1997)		
	Contact: Harvey, D.		
	G. M. Rubin-Molecular and Cell Biology		
	University of California Berkeley		
	539 LSA, Berkeley, CA 94720-3500, USA		
	Fax: 510 643 9947		
	Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu		
	Plate: 283 row: E column: 1		
	High quality sequence stop: 595.		
FEATURES			
Source	Location/Qualifiers		
	1..653		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="LD28349"		
	/clone_lib="LD Drosophila melanogaster embryo pot2"		
	/sex="male and female"		
	/dev_stage="0 to 24 hours mixed stage embryonic"		
	/lab_host="XLI Blue"		
	/note="Organ: embryo. Vector: POT2. Site-1: EcoRI; Site-2		
	XhoI; Sized fractionated cDNAs were directly ligated into		
	POT2."		
BASE COUNT	181 a 160 c 161 g 151 t		
ORIGIN			
Query Match	54.7%; Score 412; DB 14; Length 653;		
Best Local Similarity	98.6%; Pred. No. 4e-101;		
Matches 426; Conservative	0; Mismatches 5; Indels 1; Gaps 1.		
QY	323	ttagcgtcgtcattgtcctgtagagacacatgcgaagaacagctgc-tggacagagc 381	
Db	1	TTAGCGCTCGCTATTCCTTCGTGAGAAATCGACGAAACGGCTGCTTGAGACGAGGC 60	
QY	382	atgtaacaatagcctggaagatggtgatacaagttcatcgtgaagagttccattcacctgcagcgc 441	
Db	61	AGTGAATATACGCTGAGAGATGCTGATCAAGTTTCATGCAAGAGTCATTCACGTCGACGCG 120	
QY	442	gagctcatatataatctgtgcacactgcgcggaggtgtgcgtacgaagcatccggcagcg 501	

Db	121	GACCTCAATATATATCTCCGACCTCGCGGAGAGTGGGCTACGAACGCATCCGGAGCGG	180
Qy	502	gctcgcttcgaggaagctcgtcgtgcgcgttaagtaaccttcgaagctcgtcgtatgttcac	561
Db	181	GCTCGTTCGTGAGGAGAGCTCGTGGCGCTTAAGTAACTTCCAGGAGCTGCATGAGTTGCAC	240
Qy	562	caggactggtgtgataacaccagaagaccgagctcgtgtcaagttcctcgtactcgtatgcc	621
Db	241	GAGGACTGTTGGTATACACCAAGAGAGGACCGGAGTCTCTCAAGTCTTAAGTCTTGATGCC	300
Qy	622	gattcgaacctcggaanaaatctgaccagcagtagtaccagcgtctcgagagcagatattgcac	681
Db	301	GATCTGAACCTCGGAAAAAATATTGGCACCAGTAAACAGCGCCTCGAGAGACGATATTTCAC	360
Qy	682	gcatctcaaglaaaccaacagccctcgcgcgtgtgtgtgtcgcacgaagcgcagaggg	741
Db	361	GCCATCTCAAGTAAACCAACAGCCCTCGCGGTTCTGTGTGCTGCACGACGACGCGCAGAGG	420
Qy	742	gtcgcagcataa 753	
Db	421	GTCCGCAGATAA 432	
RESULT	7		
AU004911			
LOCUS	AU004911	712 bp	mRNA EST 19-JAN-1999
DEFINITION	AU004911 Bombyx mori p50(Daizo)	Bombyx mori	CDNA clone ws20982,
ACCESSION	AU004911		mRNA sequence.
VERSION	AU004911.1	GI:4162282	
SOURCE	EST.		
ORGANISM	Bombyx mori		domestic silkworm.
REFERENCE	Eukalyota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditylsia		
AUTHORS	Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.		
TITLE	1 (bases 1 to 712)		
JOURNAL	Establishment of cDNA database of Bombyx mori		
COMMENT	Unpublished (1999)		
CONTACT	Contact: Mita K		
FEATURES	Genome Research Group		
source	National Institute of Radiological Sciences		
	Anagawa 4-9-1, Inage, Chiba 263-8555, Japan		
	Email: kmila@nirs.go.jp		
	PROJECT = 'CREST project by JST'.		
	Location/Qualifiers		
	1..712		
	/organism="Bombyx mori"		
	/strain="p50(Daizo)"		
	/db_xref="taxon:7091"		
	/clone="ws20982"		
	/clone_id="Bombyx mori p50(Daizo)"		
BASE COUNT	223 a	141 c	151 g 197 t
ORIGIN			
Query Match	24.8%	Score 186.8,	DB 106; Length 712;
Best Local Similarity	58.8%;	Pred. No. 4.2e-40;	
Matches 361; Conservative	0;	Mismatches 247;	Indels 6; Gaps 2
Db	78	AACCAATTCACCGTGTTCGTGGAAGGTAACATGATAGTACGATTAACACATTTCTGGAAC	137
Qy	116	actctgagaagtaacacgaacgacatctgctcgtcgtcgcagccgctcgaagaagtggcgca	175
Db	138	ATTTCGTGAGTTTGGAGTATCACCTT---TGTGAGGAGAGCCCGTTGAATGTGGCGAG	194
Qy	176	acgtcaacgaggttaactctcgtcgtgagctgtagtatacaagaatcccaagaagtggcgatgc	235
Db	195	ACTTTAAAGGTGCATCTTTTGGACATCATGTACAAAGATCCAGAAAAATGGCGCATGA	254

QY 236 ccttcagagttatgtacagctgacatgctgacacccacccacacacaga 295
 DB 255 CATTCCAGTCATAGCTTCTTACGATGTTGGACATGACCGAGACCTGCTCCAACTC 314
 QY 296 agctaaataaargagagcgtccatctttagcgtcgtatctgctgtagaagacatgc 355
 DB 315 CAGTAAAGCTAAAGAGAGCGATCATTTATTCAGTGCAGATACCTGCTCGTTGACACATTA 374
 QY 356 gaagaaagcgtcgtgagcaggaagcatgacaataacgctgagaggtgtgtacaaattca 415
 DB 375 TGAGAAATATATACACTCCATCCAGCACAGTTTGCAGTACTTGATGATGGTTCCGATCA 434
 QY 416 tcgaaagtcattcaccctgcaagcggactcactacataatctcgagacctgcggaag 475
 DB 435 TCCAAACAAACATTCCTTATGATGCTATTTGATAGTATCTTAAGACATCACCTTCAA 494
 QY 476 tggcgtacgaacgcacccgagcggcgtcgtcttctgagagagctgctgctgaagt 535
 DB 495 TAGTGTCACAAAGATATAAAAGAGAGCTCGTTCAAGAGAGAGAGTGTGCTCCCTGTCAT 554
 QY 536 accctcagagcgtcgtgagttgacacagacgtgtgtgataacacagagacgacgcagt 595
 DB 555 ACATTGAGGAACCTGATAGTTGCATGAGAGCTGCTAATCAACAGGATACATGCTGAAT 614
 QY 596 ---cgtcgaaggtcctgctgctgctgctgacatcgtgaacctggaacacattgagcagagt 652
 DB 615 GTCCCGCACAGTATTATGATGATGCTGATTAAGACTCTCTCAATTAACCATGAAT 674
 QY 653 accagcgtcggag 666
 DB 675 ACAAGAGAGTGAG 688
 RESULT 8
 AVA04522 689 bp mRNA EST 06-FEB-2000
 LOCUS AVA04522 Bombyx mori prothoracic gland p50 5th-instar day-4 larva
 DEFINITION Bombyx mori cdna clone p50v0117 T3, mRNA sequence.
 ACCESSION AVA04522
 VERSION AVA04522.1 GI:6908610
 KEYWORDS EST.
 SOURCE domestic silkworm.
 ORGANISM Bombyx mori
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
 ; Bombycoidea; Bombycidae; Bombyx.
 REFERENCE Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.
 TITLE Bombyx mori cdna
 JOURNAL Unpublished (2000)
 COMMENT Contact: Mita K
 Genome Research Group
 National Institute of Radiological Sciences
 Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
 Email: kmita@nirs.go.jp
 method: uni-directional, sequence direction: sequenced from T3 primer
 (5' -> 3')
 Project: Silkworm Genome Program in MAFF, and Research for the
 Future Program in JSPS, see 'Silkbase',
 <http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
 FEATURES
 source
 1. 689
 /organism="Bombyx mori"
 /strain="p50"
 /db_xref="taxon:7091"
 /clone="p50v0117"
 /clone_lib="Bombyx mori prothoracic gland p50 5th-instar
 day-4 larva"
 /sex="female/male mixed"
 /tissue_type="prothoracic gland"
 /dev_stage="5th-instar day-4 larva"
 BASE COUNT 228 a 138 c 131 g 190 t 2 others
 ORIGIN

Query Match 19.2%; Score 144.4; DB 30; Length 689;
 Best Local Similarity 57.6%; Pred. No. 1.3e-28;
 Matches 296; Conservative 0; Mismatches 213; Indels 5; Gaps 2;
 QY 158 ccgtcagaagatggtgagcgaacgtcaacgggttaaatctgctgagagctgtatcaaaagtc 217
 DB 1 CCGTTGAATATGTCGGAGATCTTAAAGCTTGCAATCTTTTGGAACTCATGTAAAGATC 60
 QY 218 ccaagaatggtggcatgc---ccttcagatgtatgtacagctgaccatgctgagctgca 275
 DB 61 CCGMAAATGAGGGCGCATGAAATTCAGTACATTCCTTCTTGACGATGTTGGACATCA 120
 QY 276 caacgcccacaacaagaagcttaaaataargagcgtccatctttagcgtcgtca 335
 DB 121 CCGAGACCTGCTCCACATCCAGTAAAGCTAATGGAGCATATTAATCAGTGCAGATA 180
 QY 336 ttgctcgtgagaaatcgtcgaacggcgtcgtgagcagggcatgtacatacgtc 395
 DB 181 CTGCTTCGTTGAACATGATGAGAAATATACACTCCATCCAGCAGAGTTGCACTTCT 240
 QY 396 ggaagagtggtacaaatcatcgaagagtcattcaacctgcagggcgacatcataata 455
 DB 241 TGATGAGTGGTCCNNMTTCACACACAACTTCATTTGCTGATGTTGATGATTA 300
 QY 456 tctgcgacctgcggaaggtggtgagcgaacgacatccgcgagcggcgtcgtctgagga 515
 DB 301 TCTTAAGACATCACCTTCAATAGTGTACCAAGAAATCAAAAGAGAGCTCGTTCAAGAGA 360
 QY 516 gagctcgtgcttaagtaacctcgaagctgacgtgacgtgacacagactgtgttat 575
 DB 361 GCAGTGTGTCCTCCCTGTCATATGAGAACTGATGAGAACTGATGAGAGAGCTGCTAAT 420
 QY 576 acacacagagcagcagcagcgt---cgtcgaaggtcctgacccatgagcctgacact 632
 DB 421 CAACAGATACATGATGAAATGTCCGCCACCATGATTAATGTTGATGATTTAGACCT 480
 QY 633 ggaacacatgacacgagtaacagcgtcgtgag 666
 DB 481 CTCTCAGATTAACGATGAATACAAAGAGAGTGAG 514
 RESULT 9
 AL583655 986 bp mRNA EST 16-FEB-2001
 LOCUS AL583655 LTI_NFL006_P12 Homo sapiens cdna clone CS0DK008VA23 5
 DEFINITION Prime, mRNA sequence.
 ACCESSION AL583655
 VERSION AL583655.1 GI:12952825
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
 TITLE Full-length cdna libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1. 986
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DK008VA23"
 /clone_lib="LTI_NFL006_P12"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cdna
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cdna was digested with Not I and

Db 510 GATCATTTTACAGGCCAAGATACATATTTTGGGAAATCTTTATTAAGAATGGAAGATGC 565

Qy 374 agcaggacatgtacatacgccttgagagatggttcaacgttcatcgaaagttcat 428

Db 570 CCGAGGTGGACTTCGCTCGTTTAAAGTGAATGGTTTGAATGATATCATCAAGAACAT 624

RESULT	13					
LOCUS	BF652168					
DEFINITION	BF652168	294 bp	mRNA	EST		20-DEC-2000
ACCESSION	B275654	MARC	3BOV Bos taurus	CDNA 5', mRNA sequence.		
VERSION	BF652168					
KEYWORDS	BF652168.1	GI:11917298				
SOURCE	EST.					
ORGANISM	COW.					

Db 178 ACCATGCTCGACCAACACTCGTCGCCAGACGTTGCCCTGTACCGCTTGATGAGACGGGTCA 237
 Oy 319 atttttagcgtcgcglatctgtctcgtgagacatgacgacaaacg 365
 Db 238 ATTTCAGTGCACAGTACGTTTTTGTAAGAAACATTGTACAGAAATGG 284

RESULT	14
AM226818	
LOCUS	AW226818 531 bp mRNA
DEFINITION	um62f04.y1 Susano mouse kidney mRna Mus musculus cDNA clone IMAGE:22685983 5' similar to TR:O15238 O15238 THYMIDINE KINASE ;
ACCESSION	mRNA sequence.
VERSION	AM226818
KEYWORDS	AM226818.1 GI:6556114
SOURCE	EST.
ORGANISM	house mouse, Mus musculus

Db	Accession	Version	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	103	ATTGAGGGCAATATTGCCAAGTGGGAAGAAGACGACATGCTGTGAGTTCTTCTCA	159	ATACCA	159				
Qy	133	aacgacatlttcgctcgtctgacccgagccgctgcgagaagtggcgcaacgtcaacgagggtaata	192						
Db	160	ACAGACGTCGAGGTGTAAATGAGCCTGTGCTCAAGTGGAGAAATGTCCATGCGCATTAAC	219						
Qy	193	ctctgtgagcgtgattgatacaagaatcccaagaagtggcgcaacgtcccttgcagattatgtc	252						
Db	220	CCTCTGAGCCCTCAGCTACATGATCCAGCCGATGGGGCCCTCACACATGCACACCTAGCTG	279						
Qy	253	aagctgaccatgcgtcgagtcgacacacgcccccaacaaagaagctcaaaaataargag	312						
Db	280	CAGCTCACCAATGCTGCGACACACACGCGCCCTCAGATGTACTGTACGTTATGCGTAATGCA	339						
Qy	313	cgctccattttacgcgtcgcgtatgtctgcgtggagaacatgcagcaaaagcgtcgtg	372						
Db	340	AGGTCAATTTCACGCGGCAAGATTAATTTTGTAGAAAACCTGTATAGAAATGGGAATG	399						
Qy	373	gagcagggacatgatacaatacgtcgaaga	401						
Db	400	CCCGAGGTTGACTACGCGAATCTGTCTGA	428						
RESULT	15								
LOCUS	BEI32294	664 bp	mRNA	EST	29-JUN-2000				
DEFINITION	db44f01.y1 Blackshear/Soares normalized Xenopus egg library Xenopus								
VERSION	BEI32294.1	GI:8579659							
KEYWORDS	EST.								
SOURCE	African clawed frog.								
ORGANISM	Xenopus laevis								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Amphibia; Batrachia; Anura; Mesodactyla; Pipidea; Pipidea;								
	Xenopodinae; Xenopus.								
	1 (bases 1 to 664)								
	Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,								
	Marlin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person								
	,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,								
	Waterson,R. and Wilson,R.								
	Washu Xenopus EST project, 1999								
	Unpublished (1999)								
	Contact: Sandy Clifton, Ph.D.								
	Washu Xenopus EST project, 1999								
	Washington University School of Medicine								
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA								
	Tel: 314 286 1800								
	Fax: 314 286 1810								
	Email: est@wustl.edu								
	Library constructed by Bento Soares and M. Fatima Bonaldo								
	(University of Iowa). DNA Sequencing by: Washington University								
	Genome Sequencing Center								
	Clone distribution: Xenopus clones from this library are available								
	through the I.M.A.G.E. Consortium/LMWL at: info@image.llnl.gov								
	Seq primer: -40RP from Gibco								
	High quality sequence stop: 474.								
FEATURES	Location/Qualifiers								
Source	1..664								

```

/organism="Xenopus laevis"
/db_xref="taxon:8135"
/clone="IMAGE:3301369"
/clone_id="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT73-Pac, Site_1: EcoRI, Site_2: NotI;
polyA-selected mRNA was prepared from unfertilized Xenopus

```

laeAis eggs. The library was constructed in the vector pTR73-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dT18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pTR73-Pac vector. The library contained approximately 7.2×10^5 recombinants, with average insert sizes of 1-1.5 kb."

Query Match	Similarity	10.3%	Score 76.8	DB 163	Length 664
Best Local	Similarity	53.1%	Pred. No. 1,4e-10		
Matches	188	Conservative	0	Mismatches 163	Indels 3
					Gaps 1
QY	74	tcgaaggcaacaatcgcgcagcgggaagaccacgatttgaaccactcgagaagatacaaga	133		
DB	297	tgcaggggaatatttcgaattggaaaaaacagctgcttgattatttttcta--atatac	353		
QY	134	acgaacattgcctgctgacagagcccgctcgagaagtggcgcaacgtcaacgggttaac	193		
DB	354	cagatcttgtaggtatttcgaaggagcctgtgctcaaaatggagaatattctctggccatamcc	413		
QY	194	tgtctgagctgtgtttacaaaagatacccaagaagtggcgccatggcccttgaagttatgca	253		
DB	414	caacttggttttaattgtrttacagatccttaacaaatggggcttttaactttggcagacgtacgtgc	473		
QY	254	cgtcagcaatgctgcagtcgacacccgcaccaacacagaagaagctaaataaiaaryagc	313		
DB	474	aactcacccattgttgacatttcacaccamaacatcatctcgccgtgttaaatgatatggaaa	533		
QY	314	gtctcaattttagcgcgtcgtatctgtctgtggagaacatcgcgaagaacgctgcgtg	373		
DB	534	ggctcaattgacagtgcaaaagtattctttgtgtgagaaactgtatcagaccgggaaaaatgac	593		
QY	374	agcagagcaatgtaacaatacagcggagagtggtatacaagtaatcgaagaagttcca	427		
DB	594	caagccttgatattatggcatttttaacacgaatgctgtttgatgatattgttmaagaaaca	647		

Search completed: October 28, 2001, 06:04:48
Job time: 7644 sec

OM of: US-09-416-579a-1 to: SPTRMBL_16:* out_format : pfs

Date: Oct 25, 2001 9:08 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODL=ifam+g2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09416579/runat.25102001_082605_18686/app_query.fasta_1.823
-DB=SPTRMBL_16 -OFMT=fastan -SUFFX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -GAPOP=10.000 -XGAPEXT=0.500
-DELDP=6.000 -DELXT=7.000 -START=1 -MATRIX=dlcosm62
-TRANS=human40.cdi -LIST=45 -DCCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NOR=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09416579_ecgn1_1.220 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-416-579a-1
Query length: 753
Database: SPTRMBL_16:*
Database sequences: 425026
Database length: 132305027
Search time (sec): 77.030000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
SP_invertebrate:09XZT6	1290.00	1919.25	4.5e-99	250	1 09xzt6 drosophila melanogaster
SP_virus:09J579	335.00	494.66	1.1e-19	235	1 09j579 fowlpox virus. orf fip15
SP_virus:055749	187.00	275.44	2.1e-07	195	1 055749 chilo iridescent virus
SP_virus:098468	162.50	239.19	2.3e-05	188	1 098468 paramecium bursaria chlo
SP_bacteria:09F744	162.00	237.03	2.6e-05	221	1 09f744 uncultured proteobacteri
SP_virus:09Y9P5	149.50	210.77	0.0003	527	1 09y9p5 ateline herpesvirus 3. t
SP_bacteria:09RX12	148.00	216.71	0.0004	207	1 09rx12 delinococcus radiodurans
SP_virus:040641	136.50	191.37	0.0038	527	1 040641 salmeline herpesvirus 2
SP_invertebrate:021233	136.00	192.28	0.0040	436	1 021233 caenorhabditis elegans
SP_bacteria:068872	135.00	188.89	0.0050	542	1 068872 myxococcus xanthus. hypc
SP_bacteria:051255	132.50	193.67	0.0072	205	1 051255 borrelia burgdorferi (ly
SP_plant:09M715	132.00	190.38	0.0082	274	1 09m715 zea mays (maltze). arabid
SP_virus:083342	132.00	182.90	0.0091	644	1 083342 murid herpesvirus 4. thy
SP_virus:081940	132.00	182.90	0.0091	644	1 081940 murid herpesvirus 4. thy
SP_human:09H715	129.50	187.21	0.0132	257	1 09h715 homo sapiens (human). co
SP_virus:066624	129.50	180.68	0.0285	274	1 066624 equine herpesvirus type
SP_plant:09M715	123.50	179.90	0.0405	213	1 09m715 zea mays (maltze). arabid
SP_bacteria:09ZB43	123.50	164.23	0.0505	1275	1 09zb43 streptococcus pyogenes.
SP_human:09U036	123.50	161.36	0.0526	1783	1 09u036 homo sapiens (human). k
SP_human:015038	123.50	161.36	0.0527	1791	1 015038 homo sapiens (human). k
SP_human:060362	123.50	159.08	0.0543	2296	1 060362 homo sapiens (human). s
SP_human:09U045	123.50	157.50	0.0555	2752	1 09u045 homo sapiens (human). s
SP_human:09U035	122.50	178.74	0.0488	205	1 09u035 delinococcus radiodurans
SP_bacteria:09RX11	122.50	175.85	0.0657	221	1 09rx11 orf virus. homologue of
SP_virus:085239	121.00	167.99	0.0733	542	1 085239 orf virus. homologue of
SP_bacteria:068872	119.00	165.24	0.1072	528	1 068872 myxococcus xanthus. hypc
SP_bacteria:09RSJ1	118.00	172.03	0.1156	205	1 09rsj1 delinococcus radiodurans
SP_bacteria:09FAY6	116.00	160.29	0.1916	557	1 09fay6 streptococcus equi subsp
SP_virus:09J215	116.00	160.29	0.1916	557	1 09j215 macaca mulatta (rhadinov
SP_virus:09J215	112.00	153.79	0.3607	189	1 09j215 macaca mulatta (rhadinov
SP_bacteria:09L730	112.00	159.86	0.3812	296	1 09l730 streptomyces ansochromog
SP_virus:069118	111.50	164.88	0.3868	153	1 069118 human herpesvirus 4 (eps
SP_bacteria:048993	111.00	155.20	0.4826	425	1 048993 mycoplasma capricolum. a
SP_bacteria:09S259	111.00	151.02	0.5573	629	1 09s259 streptomyces coelicolor.
SP_virus:09JW43	110.00	146.61	0.6457	956	1 09jw43 rat cytomagalovirus (scr
SP_human:09U040	110.00	144.18	0.6681	1262	1 09u040 homo sapiens (human). r
SP_human:09U040	109.50	138.54	0.7876	2207	1 09u040 homo sapiens (human). r
SP_invertebrate:09U0V2	109.00	155.90	0.6720	279	1 09u0v2 leishmania major. pos
SP_human:014888	109.00	144.40	0.7899	1038	1 014888 homo sapiens (human). mu
SP_plant:09LD52	109.00	144.40	0.7899	1038	1 09ld52 oryza sativa (rice). po

SP_bacteria:P70783 + 108.00 152.91 0.8310 331 1 P70783 actinoplanes tetchomy
SP_virus:08548 + 108.00 150.32 0.8618 445 1 08548 bovine herpesvirus 4.
SP_plant:09S08 + 108.00 142.77 0.9583 1054 1 09s08 oryza sativa (rice).
SP_invertebrate:09Y9M0 + 107.50 139.75 1.09 1367 1 09y9m0 drosophila melanog
SP_invertebrate:09JNV6 + 107.50 139.74 1.09 1368 1 09jnv6 drosophila melanog

seq_name: SP_invertebrate:09XZT6

seq_documentation_block:
ID 09XZT6 PRELIMINARY; PRT; 250 AA.
AC 09XZT6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE DNR PROTEIN
GN DNR OR DM-DNR OR CG5452.
OS Drosophila melanogaster. (fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.C., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson K.C., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borzova D., Botchan M.A., Boulter J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson G., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Fertler A.S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Honck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
RA Jalali B.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclod J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.W., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377006; PubMed=10446143;
RA Johansson M., Van Rompay A.R., Degreve B., Balzarini J., Karlsson A.;
RT "Cloning and characterization of the multisubstrate
RT deoxyribonucleic kinase of Drosophila melanogaster.";
RL J. Biol. Chem. 274:23814-23820(1999).
[3]
RN [3]
RP SEQUENCE FROM N.A.
RA Johansson M., Karlsson A.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN-OREGON R;
 RX MEDLINE=20158988; PubMed=10692477;
 RA Munch-Petersen B., Knecht W., Ienz C., Sondergaard L., Piskur J.;
 RT "Functional Expression of a Multisubstrate Deoxyriboonucleoside Kinase
 RT from *Drosophila melanogaster* and Its C-terminal Deletion Mutants";
 RL J. Biol. Chem. 275:6673-6679(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Munch-Petersen B., Sondergaard L., Piskur J.;
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AE003724; AAF55615.1; -
 DR EMBL: AF045610; AAD47355.2; -
 DR EMBL: AF185268; AAD56545.1; -
 DR EMBL: Y18048; CAB41881.1; -
 DR FlyBase: FBgn0022338; dmk.
 DR InterPro: IPR002624; -
 DR Pfam: PF01712; dmk; 1.
 KW Kinase.
 SQ SEQUENCE 250 AA; 29088 MW; 17C5DF197B8792DB CRC64;

alignment_scores: Quality: 1290.00 Length: 250
 Ratio: 5.202 Gaps: 0
 Percent Similarity: 99.200 Percent Identity: 98.400

alignment_block:
 US-09-416-579A-1 x Q9XZT6 ..

Align seg 1/1 to: Q9XZT6 from: 1 to: 250

1 ATGGCGAGGAGGAGCATCTGTGCCGAAAGGAGGAGCAAGTACGCCGAGG 50
 1 MetAlaGlnAlaIaSerCysAlaArgLysGlyThrLysTyrLacGlnG 17
 51 CACCGACCCCTTCACCGCTCTCATCGAGGCAACATCGGAGGAGGAGA 100
 17 ThrGlnProPheThrValLeuIleGlnGlyAsnIleGlySerGlyLys 34
 101 CCAAGTATTGAACCACTTGGAGAGTACAGAGCAAGCATTTCCCTGCTG 150
 34 hrThrTyrLeuAsnHisPheGlnLysTyrLysAsnSpleCysLeuLeu 50
 151 ACCGAGCCCGTCGAGAGAGTGGCCCAACGTCACGCGGTAATTCGCTGA 200
 51 ThrGlnProValGlnLysTyrPArgAsnValAlaGlnGlyValAsnLeuLeu 67
 201 GGTGATGTACAAGATCCCAAGAAGTGGGCCATGCCCTTTCAGAGTTATG 250
 67 uLeuMetTyrLysAspProLysTyrLysTyrAlaMetProPheGlnSerTyr 84
 251 TCAGCGTCACATGTCAGTGCAGTGCACACCGCCCAACCAAGAAGCTA 300
 84 alThrLeuThrMetLeuGlnSerHisThrAlaProThrAsnLysLysLeu 100
 301 AAAATTAAGGAGCGCTCATTTTAAAGCTGCTATTCCTGTCGAGAA 350
 101 LysIleMetGlnArgSerIlePheSerAlaArgTyrCysPheValGlnLys 117
 351 CAGGACGAGAAAGCGCTGTCGAGAGGAGGAGGAGTACATACCTGTCGAG 400
 117 mMetArgAlaGlnGlnLysSerLeuGlnGlnGlyMetTyrAsnThrLeuGln 134
 401 AGTGTCACAGTTCATGAAAGATCCATTCACCTGACGAGGAGGAGCTCATC 450
 134 lntTyrTyrLysPheIleGlnGlnLysSerIleHisValGlnAlaLysPheLeu 150
 451 AATATATCTCGACACTCGCCGAGAGTGGCGTACGAACGATCCGCGAGCG 500
 151 lIleTyrLeuArgThrSerProGlnValAlaTyrGlnArgIleArgGlnArg 167

501 GGCCTGTTCTGAGGAGAACTGCGTCCGCTTAAAGTACCTTCAGAGACTGC 550
 167 gAlaArgSerGlnGlnLysSerCysValProLeuLysTyrLeuGlnGlnLeu 184
 551 ATAGTTGGACGACGACTGGTTGATACCAAGAGACGAGGAGCTGCTGC 600
 184 lstLeuHisGlnAspTyrPheLeuIleHisGlnArgArgProGlnSerCys 200
 601 AAGTCTCTAGTCTCGATGCCGATCTGAACCTGGAACATTTGGACCGCA 650
 201 LysValLeuValLeuAspAlaSerLeuAsnLeuGlnAlaGlnGlyThrG 217
 651 GTACGACGCGCTCGAGAGACGATATTCGACCGCATCTTAAGTAAACAC 700
 217 uTyrGlnArgSerGlnSerSerIlePheAspAlaIleSerSerAsnGln 234
 701 AGCCCTCGCCGCTTGTGTCGCCGACGACGACGAGGCTGCCCGAGA 750
 234 lntProSerProValLeuValSerProSerLysArgGlnArgValAlaArg 250

seq_name: sp_virus:Q9J579
 seq_documentation_block:
 ID Q9J579 PRELIMINARY; PRT; 235 AA.
 AC Q9J579;

DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE ORF FPV151 DEOXYCYTIDINE KINASE.
 GN FPV151.
 OS Fowlpox virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Avipoxvirus.
 OX NCBI_TaxID=10261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Alfonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
 RL MEDLINE=20193820; PubMed=10729156;
 RL "The genome of fowlpox virus."
 RT J. Virol. 74:3815-3831(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Alfonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF198100; AAF44495.1; -
 DR InterPro: IPR002624; -
 DR Pfam: PF01712; dmk; 1.
 KW Kinase.
 SQ SEQUENCE 235 AA; 27678 MW; 60AC94B9B0B956D4 CRC64;

alignment_scores: Quality: 335.00 Length: 212
 Ratio: 2.359 Gaps: 5
 Percent Similarity: 66.981 Percent Identity: 36.792

alignment_block:
 US-09-416-579A-1 x Q9J579 ..

Align seg 1/1 to: Q9J579 from: 1 to: 235

67 GTCCTCATGAGGAGGAGCATGCGAGCGGAGAGACAGCATTTTGAACCA 116
 26 lIleSerIleGlnGlnLysIleSerAlaGlyLysSerThrLeuIleAsnI 42
 117 CTTGAGAGATGACAGACAGCATTTGCTGTCGACCGAGCCGTCGAGA 166
 42 eLeuSer.....AspAsnGlyTyrAsnValValGlnGlnProLeuGlnG 57
 167 AGTGCCCAACGTCACGCGGTAATCTGTCGAGCGAGTGAAGTCAAAAGT 216
 57 lntPArg.....GlyAsnAsnLeuLeuAspLysLeuTyrLysAsp 70

```

217 CCCAAGATGGGCGCATGCCCCCTTCAGAGTTATGTCACCGCTGCATGCT 266
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 Proseriatrpalatyrthrphglnserhisalaphetrprthrgrth 87
267 GCAGTGGCAGACGGCCCAACCAAGAGCTAA.....ATTAARG 310
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
87 rlysthrtyrilleasplaleuasnlysnlysglyasnilleleuc 104
311 AGCGCTCCATTTTAAAGCTCGCTATGCTTCGTGAGAACATGACGCA 360
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104 lvargservalpheseraplsytrillephalatrhalaleuhsasp 120
361 AACGCTCGCTGGACGACGACATGTACATACCTGGAGAGTGTACAA 410
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 lilelytyrilleasprapthrglutrpsanilletyrasnlglturserly 137
411 GTTCATGAGAGTCATTCACCTGACGCGGACCTCATCATATATCTGC 460
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
137 strmeththrgluphemetasprilelylileasprglylillelyleu 154
461 GCACCTGCGCGGAGGTGGCTGACGACGACATCGGCGCGGCTGCTT 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
154 ysthrserproasprilecystyltylsarphmetleuasnargalaarghis 170
511 GAGAGAGCTCGCGCGCTTAAGTACCTTCAGAGAGCTGCATGAGTTGCA 560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 glulinsnthrvallylileasprtylleuasnleuhsasprlyshl 187
561 CCAGAGACTGTTGATACACGACGACGCGGACGCTGTCAGAGCTCTAG 610
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
187 sglulysrpreu...sergluasnasnlglylilephelylvalproy 203
611 TCCTCGATGCGGATCTGACCTGGAAAACATTTGGCACCAGAGTACCAGCC 660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 alileuglulleasnlglyasprglyasprhelleasprserasnargln 219
661 TCGGAG.....AGCAGCATATTCAGCGCCATCTCA 690
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220 SerSerilleuSerasnilletyrasprhelleaser 231
seq_name: sp_virus:055749

```

```

seq_documentation_block:
ID 055749 PRELIMINARY: PRT: 195 AA.
AC 055749:
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE PUTATIVE THYMIDINE KINASE.
OS Chilo iridescent virus (CIV) (Insect iridescent virus type 6).
OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.
OX NCBI_TaxID=10488;
RN [1]
RP SEQUENCE FROM N.A.
RA Bahr U., Tidona C.A., Darai G.;
RL Virus Genes 0:0-0(1997).
DR EMBL: AF003534; AAB94460.1; -.
DR InterPro: IPR002624; -.
DR Pfam: PF01712; dmk: 1.
SQ SEQUENCE 195 AA; 22822 MW; 7F0228057F2CB85D CRC64;

alignment_scores:
Quality: 187.00 Length: 195
Ratio: 1.545 Gaps: 7
Percent Similarity: 62.051 Percent Identity: 27.652

alignment_block:
US-09-416-579A-1 x 055749 ..
Align seg 1/1 to: 055749 from: 1 to: 195

```

```

67 GTCCTCATGAGGCAACATCGGACGCGGAGAAAGACACATATTTGAACA 116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
4 lilecyslileasprlylillelelylaclaelysserthrvalthrhisar 20
117 CTTGAGAG.....TACAGAACGACATTTGCTGTCAGCCGACGCC 160
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
20 gleuylslysnleuyltylys.....Cystylrgluluproi 33
161 TCGAGAACGTGGCGCAACGTCACAGGGGTAAATCTGCTGAGAGCTGATGC 210
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
33 leasprlystrp.....ThrleuLeuproasnleuylr 43
211 AAAGATCCCAAGAGATGGCGCATGCCCTTTCAGAGTTATGTCACCGCTGAC 260
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
44 AsnAspmetlylsystrlylathrprophglnphglnvalleuphese 60
261 CATGCTGCAGTGCACACCGCCCAACAC...AAGAAGCTAAAAATTA 307
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 rglntyrasprglntyrlleuSerPhelysnspcyslysglutrvalvalv 77
308 RGGAGCGCTCCATTTTACGCGCTATGCTTCGTGAGAACATGCGCA 357
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
77 alglutrgcysprothrprthrserlysnillephethrserleuethile 93
358 CGAAGCGCTCGCTGAGACGCGCATGTACATACCGCTGAGAGAGTGTA 407
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
94 GluAsnAsnleupheasprleuSeralelileasprthr.....Ty 106
408 CAAGTTCATCGAAGAGTCCATTCACCTGACGCGGACCCATCATATATAC 457
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106 rasnAsnleuTyrgluargleuSerTyrglnvalAspHisphellelyr 123
458 TCGCACCTCGCGCGGAGTGGCTGACGACGATCCGCGCGGCGCTGT 507
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 lelysvalasprserglumetalaPheglualrglleylsyrsargsparg 139
508 TCTGAGAGAGCTGCTGCCGCTTAAGTACCTTCAGAGCTCATGAGTT 557
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 PheAlaglulinsnilleaserPheasprTyrlleuTyserleuglulnly 156
558 GCACACGAGACTGCTGATACACGACGACGACGCGCTGTCAGAGTCC 607
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 styralathrSerleuAla...ThrleuSerProserthr.....ValT 170
608 TAGTCTCGATGCCGATCTGACCTGGAACATTT 642
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 hrilleleasprglyserasnThrilleglulnval 181
seq_name: sp_virus:098468

```

```

seq_documentation_block:
ID 098468 PRELIMINARY: PRT: 188 AA.
AC 098468:
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE SIMILAR TO BACILLUS SUBTILIS 24.1 KDA PROTEIN.
GN A416R.
OS Paramesitium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96400190; PubMed=8806566;
RX Kufish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map
RT positions 182 to 258."
RL Virology 223:303-317(1996).
DR EMBL: U42580; AAC96784.1; -.
DR InterPro: IPR002624; -.
DR Pfam: PF01712; dmk: 1.
SQ SEQUENCE 188 AA; 22361 MW; 82CCTA4AF3F55F2A CRC64;

```


alignment_scores:

Quality: 148.00 Length: 188
 Ratio: 1.345 Gaps: 6
 Percent Similarity: 58.511 Percent Identity: 26.064

alignment_block:

US-09-416-579a-1 x Q9RXL2 ..

Align seg 1/1 to: Q9RXL2 from: 1 to: 207

```

67 GTCCTCATGAGGCGCAACATCGCAGCGGAGACACATATTGAAACA 116
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
3 ValAlaValSerGlySerGlySerGlySerThrLeuThrArgLeu 19
117 CTTC..GAGAGTACAGAGACGACATTGGCTGCTGACCGAGCCCGTCG 163
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
19 tleuAlaGluArgTyrGly.....LeuArgProValT 30
164 AGAAGTGGGCGCAACGTCACAGGGGTAAATCTGCTGAGCTGATGACAA 213
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
30 yrgIuProTyrAlaGluAsn.....ProTyrLeuGluAspHeterHis 44
214 GATCCCAAGAGTGGGCGCATGCCCTTTCAGAGTATGTCACGCTGACCAT 263
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
45 AspMetArgGlnTyrSerPheHisSerGlnValTyrPheLeuSerArgAr 61
264 GCTGCAGTGGCACACCGCC...CCAACCAAGAGAGTAAAAATPARGG 310
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
61 gleuGluGlnHisLeuGlyMetValThrGlyAlaArgTyrValIleGlu 78
311 AGCGCTCCATTTTTCAGCTGCTATGCTGCTGAGAGACATGACGACGA 360
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
78 spArgThrValPheGluAspAlaAsnIlePheAlaArgAsnLeuTyrGlu 94
361 AACGCTCGCTGAGAGCGGCGCATGTACAATACGCTGAGAGAGTGTACAA 410
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
95 SerGlyGlnMetGlyGluArgAspArgGlnTyrArgGlyLeuTyrGln 111
411 GTTCATCGAAGATCCATTCACCTGCGAGCGACCTCATCATATATCTCC 460
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
111 uGlyValLeuProAlaLeuArgVal...ProAspLeuLeuIleHisIle 127
461 GCACCTCGCGCGAGTGGCGTACGACGATCGGCGGCGGCTGCTGT 510
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
127 spAlaGlyLeuProThrLeuArgArgArgIleAlaLeuArgGlyArgAsp 143
511 GAGGAGAGCTGCGCGCTTAAGTACCTTCAGAGCTGCATGAGTTGCA 560
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
144 TyrGluGlnAlaIleProAspGlnTyrLeuAlaGlyLeuAsnArgLeuTyr 160
561 CCGAGCATGCTGATACACGAGAGAGAGCGGCGAGTGCACAGCTCCG 610
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
160 rAlaGlyTyrPile.....AlaAlaPheAspLeuGlyProValVal 174
611 TCCTCGATCGCGCAT 624
174 rGlyAlaSpGlyAsp 178
seq_name: sp_virus:040641
seq_documentation_block:
ID 040641 PRELIMINARY; PRT; 527 AA.
AC 040641:
DT 01-JAN-1998 (TREMBLrel. 05. Created)
DT 01-JAN-1998 (TREMBLrel. 05. Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15. Last annotation update)
DE THYMIDINE KINASE.
OS Salimidine herpesvirus 2.
OC Viruses: dsDNA viruses, no RNA stage: Herpesviridae;
OC Gammapherpesvirinae; Rhadinovirus.
OK NCBI_TaxID=10381;
RN [1]

```

RP SEQUENCE FROM N.A.

RC STRAIN-C-488;
 RX MEDLINE-98037620; PubMed-9371569;
 RA Knäppe A., Hiller C., Thürau M., Wiltmann S., Hofmann H.,
 RA Fleckenstein B., Fickenscher H.;
 RT "The superantigen-homologous viral immediate-early gene iel4/vsag in
 RT herpesvirus salimidi-transformed human T cells.";
 RL J. Virol. 71:9124-9133(1997).
 DR EMBL.Y13183; CAA73635.1; ..
 DR InterPro: IPR001889; ..
 DR Pfam: PF00693; TK_Herpes; 1.
 DR ProDom: PD001519; ?; 1.
 SO SOURCE 527 AA; 60040 MW; 44975CC07265F3F2 CRC64;

alignment_scores:

Quality: 136.50 Length: 250
 Ratio: 1.075 Gaps: 14
 Percent Similarity: 50.800 Percent Identity: 27.200

alignment_block:

US-09-416-579a-1 x 040641 ..

Align seg 1/1 to: 040641 from: 1 to: 527

```

61 TTCACGTCCTCATGAGGCGCAACATCGGCGGAGAGACCATATT 110
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
210 PhePheIlePheLeuGlnGlySerIleGlyValGlyThrThrLeuLe 226
111 G.....AACCATTCGAGAGTACAGAGACGATTTGCTGCTGA 151
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
226 uLysSerMetAsnGlyIleLeuGlyGlyAsn...ValLeuAlaPhe 242
152 CCGAGCCCGTGCAGAGTGGCGCAACGTCACAGGGGTAAATTCGCTGAG 201
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
242 IsgIuProIleAlaTyrThrPheAspValPheSer...AsnSerLeuGlu 257
202 CTGATGTACAAA..... 213
258 GluValTyrTyrLeuThrLeuProAlaValGlyArgThrSerAsn 274
214 .....GATCCCAAGAGTGGCGCATGCCCTTTCAGAGTT 247
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
274 rAlaLysLeuLeuAlaCysGlnLeuLysPheAlaSerProLeuLeuAl 291
248 ATGTCACGCTGACC...ATGCTGCAGTGCACACCGGCCCAACCAAG 294
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
291 euLysThrAlaThrAspArgLeuSerSerHis.....LysAsnSer 304
295 AAGCTAAA.....ATAAGGAGCGCTCCATTTTTCAGCG 329
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
305 LeuLeuSerSerAspMetTyrValMetPheAspArgHisProLeuSerAl 321
330 TCGCTATTGCTGCTGAGAACATCGACGAAACGCTGCTGAGACGAG 379
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
321 aThrValValPheProTyrMetHisPheGlnAsnGlyPheLeu.....S 336
380 GCATGTACAATACGCTGAGGAGGTACAGATTCACGAGAGTCCATT 429
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
336 erPheSerHisLeuIleGlnLeuTyrPheSerPheLysAlaSer..... 350
430 CACCTGCAGCGGAGCTCATCATATATTCGCGACCTCGCGGAGGTGGC 479
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
351 .....GlnGlyAspAsnIleIleLeuLeuAsnLeuAsnSerGlnGluAs 365
480 GTACAGACGATCCGCGAGCGGCTCGTTCTGAGAGACATGCTGCCGC 529
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
365 nLeuLysArgValArgLysArgAsnArgLysGlnGlnLysSerValSer 382
530 TTAGTACTTCAGGAGCTG.....CATGAGTGCACAGAGAC 567
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
382 IeGluHisIleArgLeuLeuAsnAsnCysTyrHisAlaValTyrCysAla 398

```


OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cytophacterales; Myxococcaceae; Myxococcus.
 RN NCBI_TaxID=34;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DK 101;
 RA Harris B.2., Singer M.H.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF055904; AAC82365.1;
 KW Hypothetical protein
 SQ SEQUENCE 542 AA; 57203 MW; DA7E39075A86A8BC CRC64;

alignment_scores:

Quality: 135.00 Length: 272
 Ratio: 1.107 Gaps: 19
 Percent Similarity: 44.853 Percent Identity: 30.515

alignment_block:

US-09-416-579a-1 x 068872 ..

Align seg 1/1 to: 068872 from: 1 to: 542

```

14 CATCTGTCCTCCGAAAG.....GACCAAGTACGCGGAGCAACCG 57
   ::::: |||::: ||||| ::::: |||||
44 TYRPROProGlnAlaValProGlyProCysAlaProAlaProPr 60
   ::::: |||::: ||||| ::::: |||||
58 CCCTTCACGCTCATGAGGCGCAACATGCGGACGAGCAACACGTA 107
   ||||| ||| ||| ::::: |||||
60 oargSerPro.ProProAlaArgAlaArgAlaArgSer 76
   ||||| ||| ::::: |||||
108 TTGGAACCACTTGAGAACTACAGACGACATTTGCTGTGACGAGC 157
   ||::: |||::: |||::: |||::: |||:::
77 PheArgPro..ThrcysAlaArgArgSerGlyArgArgCys...ProAla 91
   ||||| ||| ::::: |||||
158 CCGTGCAGAGATGGCGCAACGTCACAGCGGATTAATCTGCTGAGCTGATG 207
   ||||| ||| ::::: |||||
92 ProSerAlaArgAlaArgAlaArgPro..... 100
208 TACAAGATCCCAAGAGTGGGCA.....TGC...CCTTCAGAG 245
   ::::: |||::: |||::: |||::: |||:::
101 .....AlaGlyProSerAlaArgCysGlyProProAlaGly 112
   ::::: |||::: |||::: |||::: |||:::
246 TTATGTACGCTGACCATGCTGCAATGCGACACG...CCCCAACACAA 292
   ::::: |||::: |||::: |||::: |||:::
112 ySerThrProThrArgCys.....ThrProCysProAlaProPro 126
   ||||| ||| ||| ||| |||
293 AGAAGCTAAATATAGAGCGCTCATTTTACGCGCTATGCTTC 342
   ||||| ||| ||| ||| |||
127 ATGcysAlaArgAlaArgProSerAlaPro.....AlaProSerAl 139
   ||||| ||| ||| ||| |||
343 GTGAGAACATGCGAGCAACGCTCGCTGAGAGGAGCA.....TGTA 386
   ||||| ||| ::::: |||::: |||::: |||:::
139 aglyArgThrCysProThrAlaGlyProSerSerCysAlaSerGlyCys 156
   ||||| ||| ::::: |||::: |||::: |||:::
387 CAATACGCTGAGAGTGTACAGT..... 412
   ||| ||| ::::: |||::: |||::: |||:::
156 ySProThrGlyArgCysGlySerAlaProGlyProThrProHisGlyAla 172
   ||||| ||| ||| ||| |||
413 .....TCATGAGAGTCACTCACTGAGAGGAGGAGGAGGAGGAGTATC 450
   ||||| ||| ||| ||| |||
173 GluProSerProProSerGlnSerPro..SerProAlaProAlaSer... 187
   ||||| ||| ||| ||| |||
451 ATATATCTGCGACCTCGCGAGAGTGGCGTACGACGACGAGCGAGCG 500
   ::::: |||::: |||::: |||::: |||:::
188 .....SerGlyAlaArgAlaArgThrHisProArgPro 198
   ||||| ||| ||| ||| |||
501 GAGCTGTTCTGAGAGAGCTGCGCTTAACTACCTGAGAGAGCTGC 550
   ::::: |||::: |||::: |||::: |||:::
199 Ala.....ArgAla.....SerAlaAlaG1 205
   ||||| ||| ||| ||| |||
551 ATGAGTTGCACGAGACTGTTGATACACGAGAGAGAGCGGAGTGTGC 600

```

seq_name: sp_bacteria:051255

seq_documentation_block:

ID 051255 PRELIMINARY; PRT; 205 AA.

AC 051255;

DT 01-JUN-1998 (TREMBLrel. 06, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)

DE DEOXYGUANOSINE/DEOXYADENOSINE KINASE(1) SUBUNIT 2 (DCK).

GN BB0239.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RX MEDLINE=98065943; PubMed=9403685;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,

RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,

RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S., Hanson M.,

RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,

RA Ullrich T., Wathey L., McDonald L., Artlich P., Bowman C.,

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,

RA Smith H.O., Venter J.C.;

RT "Genomic sequence of a Lyme disease spirochete, Borrelia

RL Nature 390:580-586(1997).

DR EMBL: AE001134; AAC66630.1; -.

DR TIGR: BB0239; -.

DR InterPro: IPR002624; -.

DR Pfam: PF01712; dmk; 1.

SQ SEQUENCE 205 AA; 24116 MW; F25F4100DD7F1C6F CRC64;

alignment_scores:

Quality: 132.50 Length: 196
 Ratio: 1.380 Gaps: 9
 Percent Similarity: 48.980 Percent Identity: 24.490

alignment_block:

US-09-416-579a-1 x 051255 ..

Align seg 1/1 to: 051255 from: 1 to: 205

```

67 GTCTCATGAGGCAACATCGGAGCGGAGAGACGATATTGAC.. 114
   ::::: |||::: ||||| ::::: |||||
7 TleValIleGluIleuIleGlyAlaGlyThrThleuGlyAsn1 23
   ||||| ||| ||| ||| |||
115 .....CACTGAGAGTACAGAGAGAGACA 139
   ::::: |||::: |||::: |||::: |||:::
23 eleuSerLysGluLeuGluValProPheTyrSerGluLeuAsnAsp 40
   ||||| ||| ||| ||| |||
140 TTTCCTGCTGACCGAGCCGCTGAGAGAGTGGCGCAACGTCACGCGGTA 189

```

```

40 hefrLeuAla..... 43
190 AATGCTGAGCGATGATCAAGATCCCAAGAAAGTGGCCATGCTT 239
44 ...ValleusaplysphenylyLysaspLysSerArgTPrAlaPheProVa 59
240 TCAGAGTTATGTACGCTGACATGCTGACAGTGCACACCGCCCAACCA 289
59 IGLn.....IleasnPheLeu.....A 65
290 ACAAGAAGCTAAAAATAAAGGACCGCTCCATTTTACCGCTCCATTCG 339
65 snGLuArgPheLysLeuIle...LysGLyAlaPheArgTPr..... 77
340 TTCGTGGAGAACATGCGCAAGACGCTCGCTGAGCAGGCGCATGTAC.. 387
78 .....LysGLyGLyIleLeuAspArgSerIleTyrGL 88
388 .....AATACCTGGAGG 400
88 YAspCysValPheAlaSerLeuLeuAsnCysAspGLyHisIleSerAspG 105
401 AGTGTACAACTCATGAGAGATCCATT.....CACTG 435
105 LuGLuTyrLysIleTyrIleAspLeuAspAsnMetLeuGLuHisSer 121
436 CAG...GGGACCTCATCATATATATCTCGCACCTCGCCGAGGTGGCTA 482
122 GlnArgProSerLeuLeuValTyrLeuAspLysSerIleAspGLuValGL 138
483 CGAACGCTATCCGACGCGGCTCTGCTGAGAGAGCTGCGCTGTTA 532
138 nArgArgIleLysAsnArgAsnArgSerPheGLuMetAsnIleProArg 155
533 AGTACCTTCAGGAGCTGCATGTTGACACGAGACTGG 570
155 sPTyrLeuGLuIleuAsnArgLysTyrLeuLysTPr 167

-seq_name: sp_plant:09M715

seq_documentation_block:
ID 09M715 PRELIMINARY; PRT; 274 AA.
AC 09M715:
DT 01-OCT-2000 (Tremblere). 15, Created)
DT 01-OCT-2000 (Tremblere). 15, Last sequence update)
DT 01-MAR-2001 (Tremblere). 16, Last annotation update)
DE ARABINOGLACTAN PROTEIN (FRAGMENT).
GN AGP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Ponce G., Lujan R., Campos M.E., Reyes A., Nieto-Sotelo J.,
RA Feldman L.J., Cassab G.I.:
RT "The quiescent center controls the expression of three cap genes in
RT the maize root meristem."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF134579; AAF3497.1; -.
DR InterPro: IPR002965; -.
DR PRINTS: PR01217; PRICHEXTENSN.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 274 AA; 27957 MW; BFE373BB6CAB713 CRC64;

```

alignment_scores:

Quality:	132.00	Length:	259
Ratio:	1.128	Gaps:	16
Percent Similarity:	45.174	Percent Identity:	30.116

alignment_block:

US-09-416-579a-1/rev x 09M715 ..

Align seg 1/1 to: 09M715 from: 1 to: 274

```

745 CGAGCCCTGCGCGCTGCTGGCGCACACAGAACCGGCGAGGCTGTGG 696
7 ArgAlaSerSerAlaAlaThrAlaThrProSerThrSerThrAlaLgl 23
695 TTACTTGAGATGGCGTGAATATGCTCTCCG..... 662
23 Y.....ThrProThrSerLacysSerProThrAlaThrSerThrs 37
661 .....AGGCTGCTGACTCGTGGCCAAATGTTTCAGGTTCA 626
37 erThrArgThrSerSerAlaSerThrAlaProThrAlaCysProGLyThr 53
625 GATCGCATGAGGAGACTAGACCTTGCACGACTGCGGTCTCTGTGTG 576
54 SerArgGLySerArgProSerProCysSerSerThrAlaThrSerSerTh 70
575 ATCAACCAAGTCTGTGTCMACTCATGCAGCTCCTGAAGT.....ACTT 532
70 rSer...AlaProGLy.....ArgArgProArgGLyThrThrThrs 83
531 AAGCGCGACGAGCTCTCCCTCAGAAC...GAGCCGCTGGCGGATCGTT 485
83 erThrAlaThrPserSerProThrAlaSerProCysAlaSerCys... 98
484 CGTACGCGACCTCCGCGGAGGTGGCAGATATATGATGAGTCCGCTTCG 435
99 ArgGLuProThrProArg.....GlyArgProAla 108
434 AGCTAATGAGACTCTTCGATTAACCTTACACACCTCCAGCGATATGTA 385
108 A.....ProCysArgArgCysP 114
384 CATGCCCTGCTTCACAGCGCGTTTCTGCGCAGTTCACGAGCAAT 335
114 roSerProAlaProAla...ArgProThrAlaCysSerSerArgSerThr 129
334 AGCGAGCGCTAAAAATGAGCGCTCTTATTTTACCTTCTGTGTGTT 285
130 Ala.....GlySerProSerGLyProThrProCys..... 139
284 GGGGCGGTGTGCGATGCGACGATGTCAGCGTGACATTAACCTGTAAGG 235
140 .....ProSerGLnArg 144
234 CATGGCCACTTCTTGGGATCTTTGTACA.....TCAGCTCCAGCA 194
144 rGLySerProGLy.....CysThrGLyThrAlaSerArgProThr 156
193 GATTTAACCCGTTGACGTTGCGCCACTTCTGACGGGCTCGGTGACGAG 144
157 ThrAlaSerArgThrLeuThrTrpProSerSerSerAlaArgSerProp 173
143 CAATGCTGTTCTTACTTCTGCAAGTGTCAATATGCTGTGCT.... 98
173 oThrCysThrAlaThrPserAlaArgThrAlaProThrThrSerThr 190
97 .....TCCGCTGCGCGAGTGGCTCGATGAGGAGCGGTGAAGGCT 56
190 LysSerMetSerArgProGLnCysProProTPrGLyGLy.....ThrAla 204
55 GGGTGGCTTGGCGGCTACTGTGTCCTT 29
205 ThrThrProArgProAlaSerSerPro 213

```

seq_name: sp_virus:083342

seq_documentation_block:

ID Q83342 PRELIMINARY; PRT; 644 AA.
 AC Q83342;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DE 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
 DE THYMIDINE KINASE.
 OS murid herpesvirus 4.
 OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 NC NCBL_TaxID=33708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G24;
 RA MEDLINE=96213518; PubMed=8638414;
 RA Pepper S.D., Stewart J.P., Arrand J.R., Mockett M.;
 RT "Murine gammaherpesvirus-68 encodes homologues of thymidine kinase and
 RT glycoprotein H: sequence, expression, and characterization of
 RT pyrimidine kinase activity."
 RL Virology 219:475-479(1996).
 DR EMBL; X93468; CAA63755.1; -;
 DR InterPro; IPR001889; -;
 DR Pfam; PF00693; TK_Herpes; 1.
 DR Prodom; PD001519; -; 1.
 DR SEQUENCE 644 AA; 7225 MW; FDF782746B557A9 CRC64;

alignment_scores:

Quality: 132.00 Length: 275
 Ratio: 0.957 Gaps: 11
 -Percent Similarity: 50.182 Percent Identity: 22.909

alignment_block:

US-09-416-579a-1 x Q83342 ..

-Align seg 1/1 to: Q83342 from: 1 to: 644

```

22 GCCCGAAGGGGACCAAGTACGCCGAGGCG...ACCGAGCCCTTCACCGT 68
312 SerlysmetSerThrAspPheProAlaGlyProThrArgCysAlaThrVa 328
69 CCTC...ATCGAGGGCAACATCGCGAGCGGAGACCACTATTGATGAC 115
328 lleuTyrrhepspllyleAlaValaIglylysthrThAlaLeuGlna 345
116 ACTTCGAGAGTACAGAGACATTTGCTGCTGACCGAGCCCGTCGAG 165
345 laAlaAlaGluPheLeuHisAsnIleuValIleSerGluProIlePro 361
166 AAGTGGCGCAACGTACAGGGGTAATCTGCTGAGCTGATGTAC..... 210
362 TyrTrpThrSerThrPheAspLysAsnValCysGlnGlnIleTyrAspVa 378
211 .....AAGATCCCAAGAG..... 225
378 lVallysthrLysGluLysGlyLysLysHisSerLysLysValLeuGlnC 395
226 .....TGGGCATGCGCCTT.....CAGGATATGTACAG 255
395 ySglnMetAlaPheAlaGlnProPheHisAlaThrGlnHisLeuLeuArg 411
256 CTGACCATGCTGCAGTGCACACACGCCCA...ACCAACAGAGACTAA 302
412 ArgThrCysLeuGlyThrLysValSerAspGlyThrCysAsnAsnTyrVa 428
303 AAATAAGGAGCGCTCCATTTTAGCGCTGCTATGCTTCGTGAGAACCA 352
428 lleuIleAspArgHisMetValSerProThrValIlePheProCysLeuP 445
353 TGGCGAAGAACGGCTGCTGAGAGGAGGATATACATAAGCTGAGAGAG 402
445 hepheArgLeuGlyValLeuArg.....PheCysAspLeuSerLeu 459
403 TGGTACAAAGTTCATCGAAGAGTTCATTCACCTGAGGCGGACCTCATCAT 452

```

seg_name: sp_virus:041940

```

seq_documentation_block:
ID 041940 PRELIMINARY; PRT; 644 AA.
AC 041940;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE THYMIDINE KINASE.
GN GAMMAHV. ORF21.
OS murid herpesvirus 4.
OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBL_TaxID=33708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WMMS, AND G2.4;
RA Virgin H.W. IV, Latreille P., Wamsley P., Hallsworth K., Weck K.E.,
RA Dal Canto A.J., Speck S.H.;
RT "Complete sequence and genomic analysis of murine gammaherpesvirus
RT 68".
RL J. Virol. 71:5894-5904(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WMMS;
RA Latreille P., Wamsley P., Waterston R.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=G2.4;
RA MEDLINE=96213518; PubMed=8638414;
RA Pepper S.D., Stewart J.P., Arrand J.R., Mockett M.;
RT "Murine gammaherpesvirus-68 encodes homologues of thymidine kinase and
RT glycoprotein H: sequence, expression, and characterization of
RT pyrimidine kinase activity."
RL Virology 219:475-479(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=G2.4;
RA Milligan S., Efsthaliou S., Stewart J.P., Nash A.A., Davison A.J.;
RT "Genetic content of murine gammaherpesviruses."

```

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U97553; AAB6394.1; -
 DR EMBL: AF105037; AAP19286.1; -
 DR InterPro: IPR001889; -
 DR Pfam: PF00693; TK_herpes; 1.
 DR Prodom: PD001519; -; 1.
 KW Kinase.
 SQ SEQUENCE 644 AA; 72269 MW; FD3CD1646FE5562 CRC64;

alignment_scores:

Quality: 132.00 Length: 275
 Ratio: 0.957 Gaps: 11
 Percent Similarity: 50.182 Percent Identity: 22.909

alignment_block:

US-09-416-579A-1 x 041940 ..

Align seg 1/1 to: 041940 from: 1 to: 644

```

22 GCCGGAAGGGACCAAGTACCGCGAGGC..ACCCAGCCCTTCACCGT 68
   ::::: ::::: ::::: ||| ::::: |||||
312 SerlysmetSerThrAspPheProAlaGlyProThrArgCysAlaThrVa 328
   CCTC...ATCGAGGGCAACATCGCAGCGGAGAGACACAGTATTGAACC 115
   ||| ::::: ||::: ||::: ||||| ||::: ||:::
328 lleuTyrgheaspIyglIlealValGlyLysThrThralaleuGlnua 345
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
116 ACTTCGAGAAATACAGACACACATTTGGCTCGTACCGAGCCCGTGAG 165
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
345 laalaalaglupheleuHisasIleleuValIleSerGluProIlePro 361
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
166 AAGTGGCCGCAATCAACGGGGTAAATCTGCTGGAGCGTATGATAC 210
   ||| ::::: ::::: ||| ::::: ||::: ||:::
362 TyrTrpThrSerThrPheAspLysAsnValCysIleGlnIleTyraPya 378
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
211 ::::: ::::: ::::: ||| ::::: ||::: ||:::
378 lvallystrlysglulysgllysllyshiserylsylsvalleuGlnC 395
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
226 ::::: ::::: ::::: ||| ::::: ||::: ||:::
395 ysgIleMetAlaPheAlaGlnProPheHisAlaThrGlnHisleuLeuArg 411
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
256 CTGACCAATGCTGCAGTCGACACACCGCCCA..ACCAACAGAGATGAAA 302
   ||| ||| ::::: ::::: ||| ::::: ||:::
412 ArgThrCysleuGlyThrLysValSerAspLysThrCysAsnAsnTyra 428
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
303 AATAAGGAGCGCTCCATTTTACCGCTCGTATGCTTCGTGGAGAAC 352
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
428 lleuIleAspArgHisMetValSerProThrValIlePheProCysleuP 445
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
353 TGGCAGCAAGACGCTCGCTGAGCAGGCGCATGTACATACGCTGGAGAG 402
   ::::: ||| ||| ::::: ::::: ||| ::::: ||:::
445 hePheAlaGlyleuGlyValleuArg.....PheCysAspLeuLeuSerleu 459
   ::::: ||| ||| ::::: ::::: ||| ::::: ||:::
403 TGTACAGAGTTCATCAAGAGAGTCCATCACCCTGACGAGCGACATCAT 452
   ||| ::::: ::::: ||| ::::: ||::: ||:::
460 leuSerValPheSerAlaGlnLysPhe.....AspAsnValIva 472
   ::::: ::::: ::::: ||| ::::: ||::: ||:::
453 ATATCTGGCAGCTCGCCGAGAGTGGCGTACGAACGACATCGGAGCGGG 502
   ::::: ||| ||| ::::: ::::: ||| ::::: ||:::
472 lleuPheLysleuHisProLysValAlaAlaThrArgVallysgluArg 489
   ::::: ||| ||| ::::: ::::: ||| ::::: ||:::
503 CTGCTTGGAGAGAGCTGGCTGCGCTTAAGTACCTTCAGAGACTCAT 552
   ::::: ||| ||| ::::: ::::: ||| ::::: ||:::
489 lYArgAsnProGluGlyValIleAsnHisThrTyrlleuGlnleuLeuAsn 505
   ::::: ||| ||| ::::: ::::: ||| ::::: ||:::
553 GAG.....TTCACACAGAGATGGTTGATACACAGAGACAGCAGC 590
   ||| ::::: ||| ||| ::::: ||| ::::: ||:::
506 GluValleuAspAlaIlePheCysAlaIleProIleuLeuAlaIleValPr 522
   ::::: ||| ||| ::::: ||| ::::: ||:::
591 GCACTGCTGCAAGTCTAGTCTCGATGCCGATCTGAACCTGGAAAAACA 640

```

seq_name: sp_human:09H7T3

seq_documentation_block:

ID 09H7T3 PRELIMINARY; PRT; 257 AA.
 AC 09H7T3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CDNA FLJ14280 F1S, CLONE PLACE1005584, WEAKLY SIMILAR TO TRANS-ACTING
 DE TRANSCRIPTIONAL PROTEIN ICP0.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isonai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA WagaSuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niomiya K., Iwayanagi T.,
 RT "NEBO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK024342; BAW14892.1; -
 SQ SEQUENCE 257 AA; 26269 MW; B914A57EA54DF418 CRC64;

alignment_scores:

Quality: 129.50 Length: 258
 Ratio: 1.295 Gaps: 18
 Percent Similarity: 38.760 Percent Identity: 27.519

alignment_block:

US-09-416-579A-1/rev x 09H7T3 ..

Align seg 1/1 to: 09H7T3 from: 1 to: 257

```

730 TGCTGGCGGACACAGACGAGCGGCTGTGTACTTGAGATGGCG 681
   ||| ||| ::::: ||| ||||| ||| ||| |||
49 CysThrAlaGlyArgArgProAsnArgAlaSerGly.....ArgArgArg 63
   ::::: ||| ||| ::::: ||| ||||| ||| ||| |||
680 TCGAATATGCTGC...TTCGAGCGCTGTACTCGTCCCAATGTTTC 634
   ||| ||| ::::: ||| ||| ||| ||| ||| |||
63 gArgSerCysSerProAlaProThrTrpProLeuGys...CysTyPr 79
   ::::: ||| ||| ::::: ||| ||||| ||| ||| |||
633 CAGGTCAGATCGGCGATCGAGGACTAGGACCTTCGACGACTCGGCTGTC 584
   ||| ||| ::::: ||| ||||| ||| ||| |||
79 roGlnSerArg.....ProThrAlaSerAla 87
   ::::: ||| ||| ::::: ||| ||||| ||| ||| |||
583 TCTGGTGTATCAACAGTCTGTGCACTACATGCGCT.....CC 543
   ::::: ||| ||| ||||| ||| ||| ||| |||
88 AlaGly.....ProGlyAlaCysMetArgAlaSerGlyArgPr 100
   ::::: ||| ||| ::::: ||| ||||| ||| ||| |||
542 TGAAGTACTTAAACGCGACGAGCTTCCTCAGAACGAGCCC..... 500
   ::::: ||| ||| ::::: ||| ||||| ||| ||| |||

```

```
100 OHISGLYAsnThrThrAlaSerThrAlaProProAlaHisProArgProA 117
499 ....GCTGCCGATGCGTTTCGACCCACCTCCGGCAGGTGCCGAGAT 455
      ::::: ||| ||| |||||
117 rGhrProGlyGlyProAlaLeuArgProThrProArgProCysAla... 132
454 AATATGATGAGTCGCCCTGCAGTCGAATGACTCTTCATGAACCTGTATC 405
      |||||
133 .....GlyPro.....Al 135
404 CACTCCTCCAGCGTAT.....TGATACGCCCTGCTCCA..... 371
      : ||||| ||| |||
135 aProProAlaSerArgAspCysArgCysArgProArgArgTrp 152
370 .....GCGAGCGCTTTCGTCGATGTCCTCCAGCAGAC 338
152 roArgAlaGlyArgArgGlyArgArgAlaGlyAlaCysLysPro..... 166
337 AATAGCAGCGCTAAATAATGAGCGCTCCCTTATTTTAGCTTCTGTG 288
166 ..... 166
287 GTTGGGCGGTGCGACTGCAGCATGTCCAGCGTGACATACTTGAA 238
      ||||| ||||| |||||
167 .....SerCysAlaGlyAlaAlaTrpSerAla.....A 176
237 GGGCATGGCCCACTTCTTGGAATCTTGTACATCAGCTCCAGCAGATT 188
176 rGlyAlaPro.....LeuCys..... 181
187 CCCGTTGAGTTGCGCACCTTCTGCAGCGGCTCGGTCCAGCAGCAATG 138
      ||| ||::: |||||::: ||
182 .....SerTyrArgThrSerCysAlaGlySerCys 191
137 TCGTTCTTGACTTCTCGAAGTGTTCAATAATACGTGCTTCC..... 95
      | :::: ||::: :::: ||| :::: |||
191 sGlyAlaArgThrAlaProThrProAlaProThrCysAlaSerProSerA 208
94 .....CGCTGCCGATGTGCCCTCGATGAGAGCGGTGAGGGCGCTGGTGC 50
      ||||| ||| ||| :::: |||
208 laAlaAlaSerSerCysCysArgArg.....ArgArgAlaCys 220
49 .....CCTCGGCGTACTTGCTC 32
      ||| :::: |||||
221 SerSerProThrThrAlaTrpSer 228
```


Quality:	492.00	Length:	215
Ratio:	2.911	Gaps:	3
Percent Similarity:	78.605	Percent Identity:	45.116

ID	seq_name: SwissProt_39:K1TM_HUMAN	seq_documentation_block:
AC	K1TM_HUMAN	STANDARD; PRT; 232 AA.
01	NOV-1997 (Rel. 35, Created)	
01	OCT-2000 (Rel. 40, Last sequence update)	
01	OCT-2000 (Rel. 40, Last annotation update)	
DE	THYMIDINE KINASE 2, MITOCHONDRIAL (EC 2.7.1.21) (MT-TK).	
OS	Homo sapiens (Human).	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eumammalia; Primates; Hominidae; Homo	
117	CTTCGAGAGATGACAAAGACGACATTTGGCTCTGTACCGAGCCGCTGACA	166
74	eheserasn...thrtthrasvaldgluvalldlemetclnprvalleul	90
167	AGTGGCGAACGTCACGAGGGGTAAATGTCGTGGAGCGTGAATGTCACAAAT	216
90	ystrpralasnvalahisdllyhshsnprleuSerleuMetTyrhshasp	106
217	CCCAAGAGTGGGCCATGCCCTTTACAGATGATTCACGCTGACCATGCT	266
107	AlaseratgtrpglyleuthrleuglnlthrtYralnlneuthrmetle	123
267	GCAGTGGACACGCCGCCACCAACAAAGACCTAAAATARAGAGCGCT	316
123	uasplnlshtrtrpProglmetSerProvalArgleuMetclnlyrags	140
317	CCATTTTACGCGCTCGCATATGCTCTGTGGAGACATGCGACGAACGGC	366
140	erlletYserAlaArgTyrlllehehalglshsnleuTyrhrgglyly	156
367	TGCGTGAGACAGGCGCATATACATACGCTGGAGAGTGTACAACTTCAT	416
157	LysmetProgluvalasprYrAlaIleleuSerclntrpheasprtrll	173
417	CGAAGAGCCCATACCGTGGAGGCGGACGTCATATATCTGCGACCT	466
173	evalArgasnlleasprvalserValasrlylevalTyrleuAdgrthrt	190
467	CGCGGAGAGTGGCTGACGAACGACATCGGCGAGCGGCTCTGTGAGAG	516
190	hrProgluIleCysTyrGlhArgleuYllysmetArgCysarggluGluln	206
517	AGCGGCGGCGGCTTACGACCTTCAGAGACGTCGATGAGTGCACAGGA	566
207	LysvalIleerProMetGlulTyrleuIlnshAlaIlehisArgleuTyrGluln	223
567	CTGGTTGATACAC...CAGAGACGACGCGAGCTGTCACAGCGCTTATGTC	613
223	utrpLeuvalasnlyserleuthrProAlaIlaIalArprovalleuValI	240
614	TGCGATCGGCATCTGCACCTGAAAACATTTGGCAGCGAGTACGACGCTCG	663
240	legluAlaAspHisAsnIleuGlulYllysmetleuGluluthrPhegluGlulnAsp	256
664	GAGAGACGACATATTCGAGCGGCATCTGCAGTAAACCAACAGCGCT	706
257	ArgAlaArgIleleuthrProgluAsntrpLysHsIlyprto	270

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97236800; PubMed=9079672;
RA Johansson M., Karlsson A.;
RT "Cloning of the cDNA and chromosome localization of the gene for
RT human thymidine kinase 2."
RL J. Biol. Chem. 272:8454-8456(1997).
RP SEQUENCE FROM N.A., SEQUENCE OF 1-28, AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=99142705; PubMed=9989599;
RA Wang L., Munch-Petersen B., Herström Sjöberg A., Hellman U.,
RT Bergman T., Joernvall H., Eriksson S.;
RT "Human thymidine kinase 2: molecular cloning and characterisation of
RT the enzyme activity with antiviral and cytostatic nucleoside
RT substrates."
RL FEBS Lett. 443:170-174(1999).
CC -1- FUNCTION: DEOXYRIBONUCLEOSIDE KINASE THAT PHOSPHORYLATES
CC THYMIDINE, DEOXYCYTIDINE, AND DEOXYRIDINE. ALSO PHOSPHORYLATES
CC ANTI-VIRAL AND ANTI-CANCER NUCLEOSIDE ANALOGS.
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE
CC 5'-PHOSPHATE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER, PANCREAS,
CC MUSCLE AND BRAIN.
CC -1- SIMILARITY: BELONGS TO THE DCK/DCK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstatio
CC the European Bioinformatics Institute. There are no restrictions o
CC use by non-profit institutions as long as its content is in the p
CC modified and this statement is not removed. Usage by and for commu
CC entities requires a license agreement (See http://www.isb-sib.ch/annou
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U77088; AAC51167.1; -.
DR EMBL; Y10498; CAA71523.2; -.
DR MIM: 188250; -.
DR InterPro: IPR002624; -.
DR Pfam: PF01712; DCK; 1.
KW Transferase; Kinase; DNA synthesis; ATP-binding; Mitochondrion.
FT NE-BIND 24 31 ATP (POTENTIAL).
FT FTBL 1 8 VORAMP -> MGAFCORPSS (IN REF. 1).
FT FTBL 28 28 S -> G (IN REF. 1).
FT FTBL 232 AA; 27494 MW; 479C66CB57F1AC0F CRC64;
SQ SEQUENCE

```



```

208 rleuglalailehishisleuhsiglultripleuileylslyserl 225
584 GAGCAGCCGACGTGCAAGGTCTAGTCTGATGCGGATCGAACCCTG 633
225 eupheprovalalalalaprovalleuvalileglualalaspshiswet 241
634 GAAAACATTGGCAGCGAGTACCAGCGCGGAGGAGCAGACATA 675
242 gluarqmetleuglnleupheglulgnasnarqspargille 255

seq_name: SwissProt_39: DCK_MOUSE

seq_documentation_block:
ID DCK_MOUSE STANDARD; PRT; 260 AA.
AC P43346;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE KINASE (EC 2.7.1.74) (DCK).
GN DCK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014182; Pubmed=7929097;
RA Karlsson A., Johansson M., Eriksson S.;
RT "2 cloning and expression of mouse deoxycytidine kinase. Pure
RT recombinant mouse and human enzymes show differences in substrate
RT specificity.";
RL J. Biol. Chem. 269:24374-24378(1994).
CC -!- FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL
CC DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY
CC EMPLOYED AS ANTIVIRAL AND CHEMOTHERAPEUTIC AGENTS.
CC -!- CATALYTIC ACTIVITY: NTP + DEOXYCYTIDINE = NDP + CMP.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DCK/FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
DR EMBL; X77731; CAA54787.1; -.
DR MGD; MGI:102726; DCK.
DR InterPro: IPR002624; -.
DR Pfam: PF01712; DNK; 1.
DR Transferrase; Kinase; ATP-binding; Nuclear protein.
KM NP_BIND 28 35 ATP (PROBABLE).
FT SEQUENCE 260 AA; 30367 MW; 778FDDDD07407BDE CRC64;
SO

alignment_scores:
Quality: 358.50 Length: 232
Ratio: 2.374 Gaps: 5
Percent Similarity: 65.086 Percent Identity: 33.621

alignment_block:
US-09-416-579a-1 x DCK_MOUSE ..
Align seg 1/1 to: DCK_MOUSE from: 1 to: 260

19 TGTGCCGAAGGAGGACAGTACGCGGAGGAGCAGCCAGCCCTTACCGT 68
111 :|||||: :|||||: :|||||: :|||||: :|||||: :|||||: :
9 CyspserProserThrser...SerGlyThrArgIleYsIil 24
69 CCTCATGAGGAGCAACATCGGACGCGGAGGAGGAGCAGCTATTGAACCACT 118
: |||||||: |||||||: |||||||: |||||||: |||||||: |||||||:

```

```

24 eserileglulYasnllalaalaglysserThrPheValAsnIlel 41
119 TCAGAGAAGTCAAGAACGACATTTGGCTGTGACGACGCGCTCGAAG 168
41 eulysglinalasergluasPTPrGluValProGluProValAlaArg
169 TGGCGCAACGTC..... 180
58 TrpCysAsnValGlnserThrGlnGluIupheGluIuleThrThrse 74
181 .....AACGGGTAAATCTGCTGAGACTGATGTCAAGATCCAGAA 223
74 rGlnYsserGlyGlyAsnValLeuGlnmetetylGlyIulysProGluA 91
224 AGTGGCGCATGCCCTTCAGAGTTATGTACGCTGACCATGTGCGAGTGC 273
91 rGtrpserPheThrPheGlnserThrAlaCysleuSerArgIleAla 107
274 CACACCGCCCAACCAACAGAAAGCTAAAA..... 303
108 GlnleuAlaserleuAsnGlyIuleYsAspAlaGluIulysProValle 124
304 .AlAARGACCGCTCCATTTTACGCGCTGTTCCTGCTGAGAGAA 352
124 uphePheGluArgSerValTyriserAspArgTyrllePheAlaSerAsn 141
353 TGGCAGCAAGCGGCTGCGTACGAGGAGCATGTACATACGCTGAGAG 402
141 eulYrGluSerAspCysMetAsnGluThrclulrPheThrIleYrGlnAsp 157
403 TGTACAAAGTTCATGCAAGATCCATT.....CACCTGACGCGGAGCT 446
158 TrpIAspTrpmetAsnserGlnPheGlyGlnserleuIuleuAspI 174
447 CATCATATATCTGGCAGCTGCGGAGGAGTGGTACGAGCATCGGCG 496
174 ylleIleYrleuArgAlaThrProGluIulysCysleuAsnArgIleYr 191
497 AGCGGCTGCTTGTGAGAGAGTGCCTGCGGCTTAAGTACCTTCAGAG 546
191 eulArgGluArgAsnGluIulGlnIleProleuGluIulYrleuGluIulys 207
547 CTGATGATGTTGCCACGAGACTGTTATATACACGAGAGACCGCAGTC 596
208 leuIstrYrIlyshIsGluSerTrpleuIleuIshArgThrleuYstrIs 224
597 GTGC.....AAGGCTAGTCTGATGATGATGCTGAC 630
224 rPheAspTrYrleuGlnIulValProValleuThrleuAspValAsn 239

seq_name: SwissProt_39: DCK_HUMAN

seq_documentation_block:
ID DCK_HUMAN STANDARD; PRT; 260 AA.
AC P27707;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE KINASE (EC 2.7.1.74) (DCK).
GN DCK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
[1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 58-70; 119-127 AND 189-192.
RX MEDLINE=91142207; Pubmed=1996353;
RA Chottiner E.G., Shewach D.S., Datta N.S., Ashcraft E., Gribbin D.,
RA Ginsburg D., Fox I.H., Mitchell B.S.;
RT "Cloning and expression of human deoxycytidine kinase cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:1531-1535(1991).
[2]
PARTIAL SEQUENCE, AND CHARACTERIZATION.
RP

```

```

RX MEDLINE-91192170; PubMed-2013338;
RA Eriksson S., Cederlund E., Bergman T., Joernvall H., Bohman C.;
RT "Characterization of human deoxycytidine kinase. Correlation with
RT CDNA sequences.";
RL FEBS Lett. 280:363-366(1991).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE-98004502; PubMed-9342341;
RA Johansson M., Brismar S., Karlsson A.;
RT "Human deoxycytidine kinase is located in the cell nucleus.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11941-11945(1997).
CC -1- FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL
CC DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY
CC EMPLOYED AS ANTIVIRAL AND CHEMOTHERAPEUTIC AGENTS.
CC -1- CATALYTIC ACTIVITY: NTP + DEOXYCYTIDINE -> NDP + CMP.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE DCK/DCK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M60527; AAA35752.1; -
DR PIR: A38585; A38585.
DR PIR: S14321; S14321.
DR MIM: 125450; -
DR InterPro: IPR002624; -
DR Pfam: PF01712; dmk.1.
KW Transferase; Kinase; ATP-binding; Nuclear protein.
-PT NP_BIND 28 35
-PT SEQUENCE 260 AA; 30518 MW; 626B9D2D6BED8DBC CRC64;
SQ

```

alignment_scores:

Quality:	356.50	Length:	247
Ratio:	2.214	Gaps:	5
Percent Similarity:	65.182	Percent Identity:	30.769

alignment_block:

US-09-416-579A-1 x DCK_HUMAN ..

Align seg 1/1 to: DCK_HUMAN from: 1 to: 260

```

43 GCCGAGGACCCACCCCTTCACCGCTCTCATCGAGGACCAACATCGGACG 92
   :::::::::::::::::::: ::::::::::::::::::::
16 SerGluGlyThrArgIleLysLysIleSerIleGluGlyAsnIleAlaI 32
   :::::::::::::::::::: ::::::::::::::::::::
93 CGGGAAGACCGATATTGGAACCACTTCGAGAGTACAGACGACATTT 142
   :::::::::::::::::::: ::::::::::::::::::::
32 aglyLysSerThrPheValaIleIleLysGlnLeuGlyGluAspTrpG 49
   :::::::::::::::::::: ::::::::::::::::::::
143 GCCGCTACGAGCGCCGTCGAGAGAGTGGCGGCAACGTC..... 180
   :::::::::::::::::::: ::::::::::::::::::::
49 IuValValProGluProValaIlaArgTrpCysAsnValGlnSerThrGln 65
   :::::::::::::::::::: ::::::::::::::::::::
181 ..... AACGGGGAATTCGCT 197
   ::::::::::::::::::::
66 AspGluPheGluGluLeuThrMetSerGlnLysAsnGlyGlnValLe 82
   ::::::::::::::::::::
198 GGAGCTGATGTACAAAGATCCCAAGAGTGGCATGCCCTTTAGAGATT 247
   ::::::::::::::::::::
82 uGlnMetMetLysGluLysProGluArgTrpSerPheThrPheGlnThr 99
   ::::::::::::::::::::
248 ATGTACACCTGACCATGCTGAGTGGACGACGCGCCCAACCAACAGAG 297
   ::::::::::::::::::::
99 yRlaIaCysLeuSerArgIleArgIleArgIleValaIleSerLeuAsnGlyLys 115
   ::::::::::::::::::::
298 CTAAAA..... ATAAAGGAGCGCTCCATTCTTAG 326

```

```

116 LeuLysAspAlaGluLysProValLeuPhePheGluArgSerValTyrLys 132
   ::::::::::::::::::::
327 CGCTCGCTATTGCTTGTGGAGAACATCGGACGAAGAGCGCTGGAGAC 376
   ::::::::::::::::::::
132 RasArgTyrIlePheIleAsnSerAsnLeuTyrGlnSerGlyCysMetAsnG 149
   ::::::::::::::::::::
377 AGGCAATGATACATACGCTGAGAGAGTGTACAGATTCATCCAGAGTCC 426
   ::::::::::::::::::::
149 IuThrGluTrpThrIleTyrGlnAspTrpHisAspTrpMetAsnGln 165
   ::::::::::::::::::::
427 ATP..... CACCTGAGCGGACGACATCATATATCTCGACACCTCGCC 470
   ::::::::::::::::::::
166 PheGlyGlnSerLeuLeuAsnAspGlyIleIleTyrLeuGlnAlaIleThr 182
   ::::::::::::::::::::
471 GGAGTGGCGGTGAGAACGATCCGCGGCGGCGCTGTTCGAGAGAGCT 520
   ::::::::::::::::::::
182 OGluTrpCysLeuHisArgIleTyrLeuArgIleArgAsnGluGlnGln 199
   ::::::::::::::::::::
521 GCGTGGCGGTATAGTACCTTCAGAGCTGATGAGTTCAGACGAGACGTG 570
   ::::::::::::::::::::
199 IyIleProLeuGluTyrLeuGluLysLeuHisTyrLysHisGluSerTrp 215
   ::::::::::::::::::::
571 TTGATACACGAGACGACGACG..... CAGTCGTGCA 602
   ::::::::::::::::::::
216 LeuLeuHisArgThrLeuLysThrAsnPheAspTyrLeuGlnGluValPr 232
   ::::::::::::::::::::
603 GGTCTAGTCTCTGATGCGGATCTGAACCTGGAAGAACATTCGACCGAGT 652
   ::::::::::::::::::::
232 OIleuThrLeuAspValAsn..... GluAspPheLysAspLysGT 246
   ::::::::::::::::::::
653 ACCAGCGCTCGGAGACGACATATTCGACGCGCATCTCAAGT 693
   ::::::::::::::::::::
246 yGluSerLeuValGluLysValLysGluPheLeuSerThr 259
   ::::::::::::::::::::
seq_name: Swissprot_39:DCK_RAT

```

seq_documentation_block:

ID	DCK_RAT	STANDARD:	PRT:	260 AA.
AC	P48769;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	DEOXYCYTIDINE KINASE (EC 2.7.1.74) (DCK).			
CN	DCK.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-95121928; PubMed-7821805;			
RA	Stegmann A.P., Honder M.W., Willeme R., Landegent J.E.;			
RT	"Cloning of the Dck gene encoding rat deoxycytidine kinase.";			
RL	Gene 150:351-354(1994).			
CC	-1- FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL			
CC	DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY			
CC	EMPLOYED AS ANTIVIRAL AND CHEMOTHERAPEUTIC AGENTS.			
CC	-1- CATALYTIC ACTIVITY: NTP + DEOXYCYTIDINE -> NDP + CMP.			
CC	-1- SUBUNIT: HOMODIMER.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE DCK/DCK FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: L33894; AAA65098.1; -			
DR	InterPro: IPR002624; -			

Percent Similarity: 63.303 Percent Identity: 34.404

Alignment_block:

US-09-416-579A-1 x DGK_MOUSE

Align seg 1/1 to: DGK_MOUSE from: 1 to: 277

```

46 GAGGGACCCAGCCCTTACCCGCTTCATGAGGCAACATCGGACGG 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 Aspglygluylproargarglyleucylleuylasnllealvalgl 50
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96 GAAGACCCAGCTATTGAACCACTTCGAGAGTAACAGACACATTGGC 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 ylyserthrhevallyleuileumetlysthnisprodlutrglnv 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 TGCGTACCCAGCCGCTCGAGAGTGGCGCAACGTCACGGGGTA..... 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 alalathgluproilealalathtrpglnasnlleglnalalaglala 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 .....AATGCTGGAGCTGATGTA 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 glulyasrglythrserlyargleuylasnlleuileuglumetly 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 CAAGATCCCAAGAGTGGCCGATGCCCTTCGAGAGTATGTCACGGTGA 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
100 tglngluproalaargtrpserlythrphelthrleuSerphenetcs 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 CCATGCTGCTGACAGTGCACACGCCCA.....ACC 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 earglyleuylvalglnglneugluprolleproglyargleuileu 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 AACAGAGCTAAATAATARGAGCGCTTCATTTTACGCTCGCTATG 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 glulyservalargvalrhegluargservalyltysersasparglytl 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 CTTCGGGAGACATGCGCAGAAAGCGTCCGTGAGAGGAGCATGTACA 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150 ephelalaysanleuphegluasnglyserleuSerprlleglutprh 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 ATAGCGTGGAGAGTGTACAAAGTTATC.....GAAGAGTCCATTCAC 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 lslletyglslasprtnlsserpheluleuaglncgluphealasnarg 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 CTGCGAGCGGACATCATATATATGCGGACCTCGCGGAGGAGCGCA 482
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 leuileuileuileuglyrheleuyleuileuileuileuileuileu 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 CGAAGCATCGCGAGCGGCTGCTTCGAGAGAGTGGCTGCGCTTA 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 tgluargleuylrglnargdasparglucylulysgluileuileu 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533 AGTACCTTGAAGAGTGTACAGTGTGACACAGAGTGGTGTATA..... 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 latytleuileuileuileuileuileuileuileuileuileuileu 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
577 .....CACAGAGAGCGAGCGTGTGCAAGCTGCTGCTGCTGCT 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 thlthlyleuileuileuileuileuileuileuileuileuileu 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
615 CGAT 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 uasp 251

```

seq_name: SwissProt_39:DGK_HUMAN

seq_documentation_block:

```

ID   DGK_HUMAN          STANDARD;          PRT:   277 AA.
AC   Q16854; Q16759; P78532;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DEOXYGUANOSINE KINASE, MITOCHONDRIAL PRECURSOR (EC 2.7.1.113) (DGK).
GN   DGK OR DGK.

```

```

OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96293511; PubMed=8692979;
RA   Johansson M., Karlsson A.;
RT   "Cloning and expression of human deoxyguanosine kinase cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:7258-7262(1996).
RN   [2]
RP   SEQUENCE OF 18-277 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96314545; PubMed=8706825;
RA   Wang L., Hellman U., Eriksson S.;
RT   "Cloning and expression of human mitochondrial deoxyguanosine kinase
RT   cDNA.";
RL   FEBS Lett. 390:39-43(1996).
RN   [3]
RP   SEQUENCE OF 1-47 FROM N.A.
RA   Stegmann A.P.A., Mitchell B.S.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC   - FUNCTION: REQUIRED FOR THE PHOSPHORYLATION OF SEVERAL
CC   DEOXYRIBONUCLEOSIDES AND CERTAIN NUCLEOSIDE ANALOGS WIDELY
CC   EMPLOYED AS ANTIVIRAL AND CHEMOTHERAPEUTIC AGENTS.
CC   - CATALYTIC ACTIVITY: ATP + DEOXYGUANOSINE -> ADP + DGMP.
CC   - SUBUNIT: HOMODIMER.
CC   - SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   - ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4 AND 5;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   - TISSUE SPECIFICITY: UBQUITOUS. HIGHEST EXPRESSION IN MUSCLE,
CC   BRAIN LIVER, AND LYMPHOID TISSUES.
CC   - SIMILARITY: BELONGS TO THE DGK/DGK FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U41668; AAC50624.1; -.
DR   EMBL; X97386; CAAG6054.1; -.
DR   EMBL; U81499; AAB39858.1; -.
DR   MIM; 601465; -.
DR   InterPro; IPR002624; -.
DR   Pfam; PF01712; dnr; 1.
KW   Transferase; Kinase; ATP-binding; Mitochondrion; Transic peptide;
KW   Alternative splicing.
FT   TRANSIT         1..39
FT   CHAIN           40..277
FT   N_BIND          45..52
FT   N_BIND          48..85
FT   VARSPLIC        149..236
FT   VARSPLIC        47..47
FT   VARSPLIC        83..83
FT   CONFLICT        159..159
FT   CONFLICT        212..212
FT   CONFLICT        212..212
SQ   SEQUENCE 277 AA; 32069 MW; DIF8AA4C15224EE2 CRC64;

```

alignment_scores:

```

Quality: 340.00 Length: 233
Ratio: 2.329 Gaps: 5
Percent Similarity: 62.661 Percent Identity: 33.047

```

alignment_block:

US-09-416-579A-1 x DGK_HUMAN

Align seg 1/1 to: DGK_HUMAN from: 1 to: 277

[illegible]

```

CC      -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE
CC      5'-PHOSPHATE.
CC      -1- SIMILARITY: BELONGS TO THE DCK/DCK FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announcement/
CC      or send an email to license@isb.ch).
CC      -----
DR      EMBL; M75136; AAA88186.1; -
DR      EMBL; M75136; AAA88108.1; -
DR      PIR; J01336; KIBIEIC.
DR      InterPro: IPR002624; -
DR      Pfam; PF01712; DNK; 1.
DR      Transferase; kinase; DNA synthesis; ATP-binding.
KW      NP_BIND
FT      23
SE      228 AA; 25642 MW; 960026878787280B CRC64;

```

```

alignment_block:
US-09-416-579A-1 x KITH_HSVI1      ..
Align seg 1/1   to: KITH_HSVI1      from: 1   to: 228

```

```

46 GAGGGCAACCCGACCTTCACCCGCTTC.....AT 74
   |||||
5 GluGlyProThrProAlaArgPheLeuProGluGlyLeuValPheGlyVal 21
   |||||

75 CGAGGGCAACATCGGACGGGGAAGACCACTATTGAAACACTTCGAGA 124
   |||||
21 GlnGlyAsnIleGlyCysGlyLysSerThrLeuValLysAlaLeuMetG 38
   |||||

125 AG.....TACAAAGAACGAACTTGGCTCTCAGCAGCGCCGCTGAGAG 168
   |||||
38 LuArgValAlaGlySerGlyValAsnValAlaGluGluProValAspGln 54
   |||||

169 TGGCGCAACGTCAACGGGTAATTCGTGGAGCTGATGTCAAAGATCC 218
   |||||
55 TrpValAsnHisAsnGlyLysAsnTyrLeuGluLeuSerTyrTyrAspPr 71
   |||||

219 CAAGAAGCGGGCATGGCCCTTCAG.....AGTTATGCA 253
   |||||
71 OThGcIlyTrAlaValAlaProPheGlnAsnLeuValPheAspSerTyrAla 88
   |||||

254 CGCTGACCATCTGTCAGTTCGACACCGCCCAACCAACAGAACTAA 303
   |||||
88 snValGlnArgLeuGlnAsnPro.....Asp 96
   |||||

304 ATAAGAGAGCGCTCCATTTTACGGCTGC.....TATGCTTCGTGA 347
   |||||
97 IleMetGluArgSerProMetSerAlaThrArgValPheCysAlaValAs 113
   |||||

348 GAACATCGGACGA.....AACGGCTCGGAGACAGGCATTAACAATA 391
   |||||
113 nGlySerArgGlyValIleProAlaThrAlaLeuProGluMetAlaAla 130
   |||||

392 CGCTGACGAGATGGTACAAATTCGATGGAAGATGCATTCACCTGACGG 441
   |||||
130 rGgGlyAlaAlaValMetArgThrIle.....AlaThrArg 141
   |||||

442 GACATCATCATATATCTGCGACACCTCCGAGAGTGGCGGTACGAACT 491
   |||||
142 ProValPheValTyrLeuGluLeuProProGlnGluCysLeuArgMetG 158
   |||||

492 CGGAGACGGGCTCGTCTGAGAGAGCTGCGCCGCTTAAGTACCTTC 541
   |||||
158 TArgTArgTArgArgThrThrGlyGluAlaGlyLeuAspTyrGluAla 175
   |||||

542 AGGACATCATGATTCGACAGACGAGATGGTGT 573
   |||||
175 rGlyLeuHisGluArgTyrGluAlaThrPhe 185
   |||||

seq_name: SwissProt_39:KITN_HSVSA

seq_documentation_block:
ID KITN_HSVSA STANDARD: PRT: 527 AA.
AC P21293:
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THYMIDINE KINASE (EC 2.7.1.21).
GN 21 OR TK.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
CC
CX NCBI_TaxID=10383;
RN [1]
RP
RX MEDLINE=90063548; PubMed=2555434;
RT Honess R.W., Craxton M.A., Williams L., Gompels U.A.;
  "A comparative analysis of the sequence of the thymidine kinase gene

```

RT of a gammaherpesvirus, herpesvirus saimiri.";
 RL J. Gen. Virol. 70:3003-3013(1989).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92333688; PubMed=1321287;
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Bisstinger B.,
 RA Newman C., Wilmann S., Craxton M.A., Coleman H., Fleckenstein B.,
 RA Honess R.W.;
 RT "Primary structure of the herpesvirus saimiri genome.";
 RL J. Virol. 66:5047-5058(1992).
 CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE
 CC 5'-PHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES THYMIDINE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X64346; CAA45643.1;
 DR EMBL: D00543; BAA00432.1;
 DR PIR: A33374; KIBBS.
 DR InterPro: IPR00189; TK_herpes; 1.
 DR Pfam: PF00693; TK_herpes; 1.
 KW Transferase: Kinase; DNA synthesis; ATP-binding.
 FT NP_BIND 216 223 ATP (BY SIMILARITY).
 SQ SEQUENCE 527 AA: 59806 MW; 4EBE3A3A178FAD7D CAC64;

alignment_scores:
 Quality: 134.50 Length: 257
 Ratio: 1.027 Gaps: 13
 Percent Similarity: 50.973 Percent Identity: 25.292

alignment_block:
 US-09-416-579A-1 x KITH_HSVSA ..

Align seg 1/1 to: KITH_HSVSA from: 1 to: 527

```

37 AASTAGCGGAGGAGCCGACCCCTTACCGCTTCATGAGGCGACACT 86
||||| : : : : : ||| : : : : : ||| : : : : : |||
202 LysHisGlnAspProValProAlaPhePheIlePheLeuGlnUlySerI 218
87 CGGACGCGGAGGAGCAGCATTTG.....AACCACTTGAGAGAAGT 127
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 egiyValGlyLysThrThrLeuLeuLysSerMetAsnGlyIleLeuGly 235
128 ACAAGAACGACATTTGCTGCTGACCGCGCTCGAGAGTGGCGCAAC 177
||||| : : : : : ||| : : : : : ||| : : : : : |||
235 LysAsn...ValLeuAlaPheHisGluProIleAlaTyrTrpThrAsp 250
178 GTCAAGGGGTAATCTGCTGGAGCTGAGTACAA..... 213
||||| : : : : : ||||| : : : : : ||||| : : : : : |||
251 ValPheSer...AsnSerLeuGlnGluValTyrLysLeuThrLeuProAl 266
214 .....GATCCCAAGA 223
266 alyValGlyArgThrSerAsnSerAlaLysLeuAlaCysGlnLeu 283
224 AGTGGCCATGCCCTTTCAGAGTTATGACGCTGACCATGCTGACGTG 273
||||| ||||| : : : : : ||| ||| |||
283 yspHeLaserProLeuAlaLeuLysThrAlaThr.....AspArg 297
274 CAACGCGCCCAACCAAGAAGCTAAAA.....ATAAR 308
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
298 LeuSerSerProLysAsnSerLeuSerSerAspMetTrpValMetPh 314
309 GGAGCGCTCATTTTAAAGCTGCTATGCTTGGAGAGAACATCGCAG 358
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
314 eAspArgHisProLeuSerAlaThrValValPheProTyrMetHisPheg 331

```

```

359 GAACGCGCTGCTGAGCAGGCGCATGATACAAATACGCTGAGAGTAC 408
: : : : : ||| : : : : : : : : : : : : : : : : |||
331 LysAsnGlyPheLeu.....SerPheSerHisLeuIleGlnLeuTrpSer 345
409 AAGTTCATGAGAGAGTCCATTCACCTGACGAGCGGACCTCATATATCT 458
: : : : : : : : : : : : : : : : : : : : : : : |||
346 SerPheLysAlaSer.....ArgGlyAspAsnIleIleLeu 358
459 GGGACGCTCGCGGAGGAGTGGCTACGACGATCCGCGAGCGGCTGCT 508
: : : : : : : : : : : : : : : : : : : : : : : |||
358 ValnLeuAsnSerGlnGluAsnLeuLysArgValLysLysArgAsnArgL 375
509 CTGAGAGAGAGTGGCTGCGCTTAACTTACCTTCAGAGCTG..... 549
: : : : : ||| : : : : : : : : : : : : : : : : |||
375 ysgLncLysSerValSerIleGlnHisIleArgLeuLeuAsnAsnGly 391
550 ..CATGAGTTGACACGAGACTGTTGATACACAGAGA...CGACCGCA 593
||| : : : : : ||||| : : : : : ||||| : : : : : |||
392 TyrHisAlaValTyrCysAlaTrpLeuLeuValGlnAsnPheThrProG 408
594 GTCGTGCAAGTCTTACCTTCGATGCC.....GATCTGAC. 630
: : : : : : : : : : : : : : : : : : : : : : : |||
408 uGluLeuValGluValCysPheAsnAlaLysHisIleThrAspLeuSer 425
631 .....CTGAAACATTGGCAGCAGCAGTACGAGCGC 660
425 eSerLysProSerPheLeuAlaLysHisValSerThrGlu...AspMet 440
661 TCGGAGACGACGATTCGAC 681
: : : : : ||||| : : : : : ||||| : : : : : |||||
441 LeuLysSerSerIlePheAsn 447

```

seq_name: Swissprot_39:DGK2_LMAC

seq_documentation_block:
 ID DGK2_LMAC STANDARD; PRT; 223 AA.
 AC 059484;

```

DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DEOXYGUANOSINE KINASE (EC 2.7.1.113) (DGK KINASE) (DGK)
DE (DEOXYNUCLEOSIDE KINASE COMPLEX I F-COMPONENT).
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11506 / R-26;
RX MEDLINE=95204449; PubMed=7896798;
RA Ma G.-T., Hong Y.S., Ives D.H.;
RT "Cloning and expression of the heterodimeric deoxyguanosine
  kinase/deoxyadenosine kinase of Lactobacillus acidophilus R-26.";
  J. Biol. Chem. 270:6595-6601(1995).
  [2]
  RP SEQUENCE OF 1-17.
  RC STRAIN=ATCC 11506 / R-26;
  RX MEDLINE=94227067; PubMed=8172906;
  RA Ikeda S., Ma G.-T., Ives D.H.;
  RT "Heterodimeric deoxynucleoside kinases of Lactobacillus acidophilus
  R-26: functional assignment of subunits using limited proteolysis
  controlled by end-product inhibitors.";
  Biochemistry 33:5328-5334(1994).
  RL -1- FUNCTION: DGK/DK PLAYS AN ESSENTIAL ROLE IN GENERATING THE
  DEOXYRIBONUCLEOTIDE PRECURSORS, DCTP AND DATP, FOR DNA METABOLISM.
  CC -1- CATALYTIC ACTIVITY: ATP + DEOXYGUANOSINE = ADP + GMP.
  CC -1- SUBUNIT: HETERODIMER OF A DEOXYADENOSINE (DAK) AND A
  CC DEOXYGUANOSINE KINASE (DGK).
  CC -1- SIMILARITY: BELONGS TO THE DGK/DK FAMILY.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```



```

304 .....ATARGAGCGCTCCATTATTACGCT 330
77 lephgualaglyglylyphevalglinsaprgserlletygluasp 93
331 CGCATTCGCTTCGGGAGAACCGGCAAGCGCGCTGGAGCGGG 380
94 thigllylephalalysmethalsalaspplslythmetserlysva 110
381 CATTCACAAATACGCTGGAGAGTGTACAGTTTCATCGAGAGTCCATTC 430
110 lasptryllysthrtyrthrserleupheglualamelvalmethrprot 127
431 ACCTG...CAGCGGACCTCATCATATATCTGGCAGCTGGCGAGGTG 477
127 yrpheprhnsproaspvalleuiletyrleuenglaspplleuasn 143
478 GCGTACGACGACCATCGGAGCGGCTGCTTCGAGGAGAGCTCGTGC 527
144 lileuasnalgilegluenglarglyargluleucluleuenglntprse 160
528 GCTTAAGTACCTTCAGGACGTGATGATTCACAGGAGTGTGTATAC 577
160 rargserlytrpplugluumethlsthargtyrgluasntrpileserg 177
578 ACCGAGACGACCGGACGTGTCAGAGTCCATGCTCGATGCCGAT... 624
177 lypheasnalaicysprovalleuileuarglyleuasptryrleu 193
625 CTGAACCTCGAAGAACATTCGACGAGTACGCGCTCGGAGACGACAT 674
194 leuasnasppluasn.....serlleuasnll 203
675 ATTCGACGCGCATTCACAT 693
203 evalaspplnlealaser 209

seq_name: SwissProt_39:KITH_EBV
seq_documentation_block:
ID KITH_EBV STANDARD; PRT; 607 AA.
AC P03177;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THYMIDINE KINASE (EC 2.7.1.21).
GN TK OR BXLFL.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85035713; PubMed=6092825;
RA Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;
RA Little E., Zentzen J., McBride A.A., Soerensen E.T., Powell K.L.,
RA Walsh-Arrand J.E., Arrand J.R.;
RT "Identification of an Epstein-Barr virus-coded thymidine kinase.";
RL EMO J. 5:1959-1966(1986).
CC -1- CATALYTIC ACTIVITY: ATP + THYMIDINE = ADP + THYMIDINE
CC 5- PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES THYMIDINE KINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; V01555; CA24799.1; -.
DR PIR; A00615; KIBETE.
DR InterPro; IPR001889; -.
DR Pfam; PF00693; TK_herpes; 1.
KW Transferase, kinase, DNA synthesis: Early protein; ATP-binding.
FT NP_BIND 291 298 ATP (PROBABLE).
FT SEQUENCE 607 AA; 67193 MW; 97A4CCDB598A09P1 CRC64;

```

```

alignment_scores:
  Quality: 121.50      Length: 262
  Ratio: 0.927        Gaps: 12
  Percent Similarity: 50.000  Percent Identity: 22.137

```

```
alignment_block:
US-09-416-579A-1 x KITH_EBV ..

```

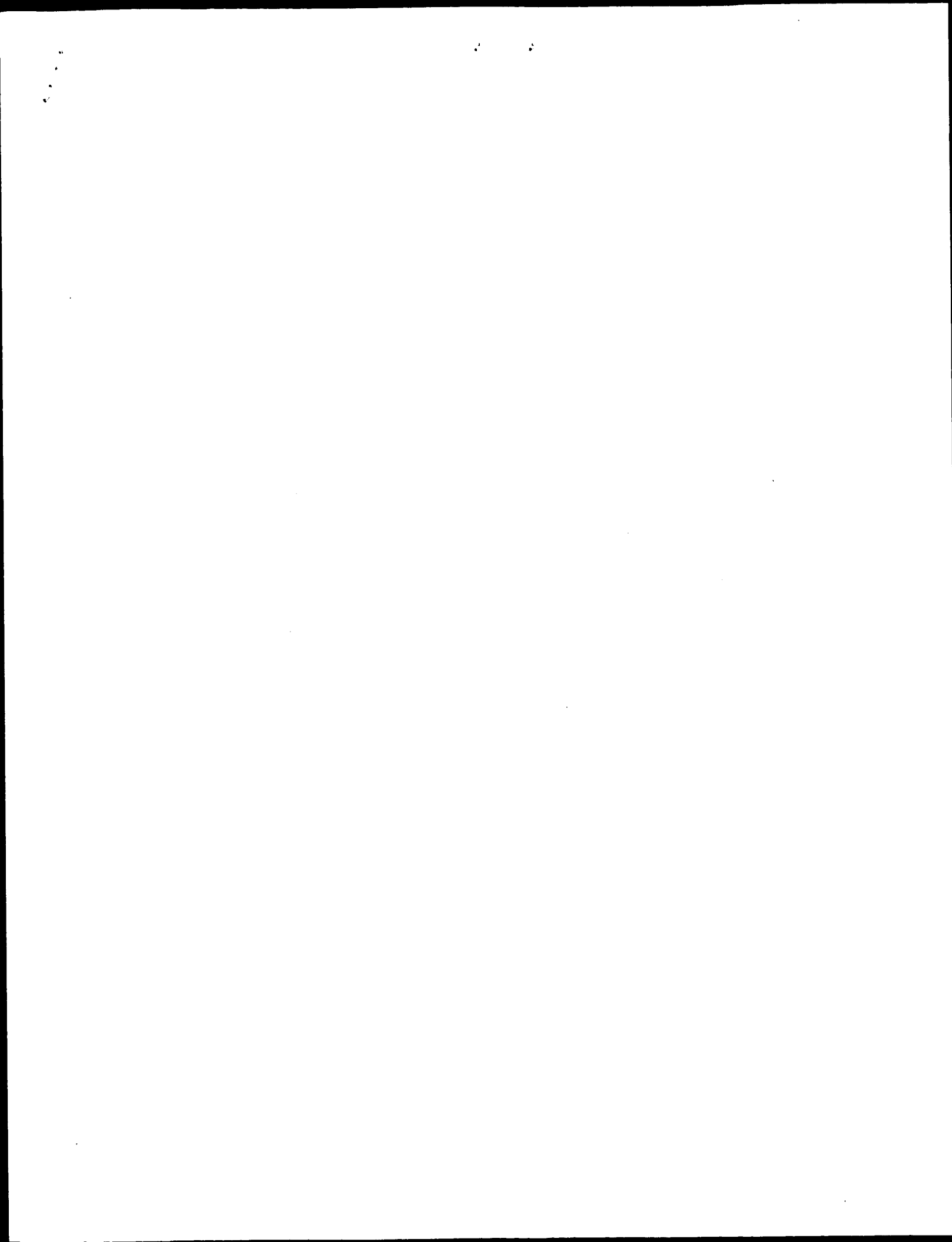
```
Align seg 1/1 to: KITH_EBV from: 1 to: 607

```

```

64 ACCGTCCTCATCGAGGCAACATCGGACGCGAAGACCGATATTGAA 113
   ::::::::::::::::::::||| ||| ||| ||| ||| ||| |||
286 Serleupheleuenglulalaprogllyvalglyllysthrmetleuasn 302
114 CCAGCTCGAGAACTACAGACGACATTCGCGGAGGAGGAGGAGGAGG 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 nhlsleuasnalaivalphealyspleuthrilevalvalprogluprom 319
161 TCGAGAGTGGCGCGCAACGTCAACGCGGTAAATCTGCTGGAGCTGATAC 210
   :::: || :::: ||| ||| :::: ||| |||
319 etarglytrpnrhnsvaltyr...gluasnalailelslamethis 334
211 AAGAT..... 216
335 Lysasnvalthrarglaalarghislglarggluaspthrseralagluva 351
217 .....CCCAAGAGTGGGCGCATGCCCTTCAGAGTATATGTC.... 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 lileualacysglmetlyserhthrprophargvalleualaser 368
253 .....ACGCTGACCATGCTGACGTG.....CACACCGCCCA 285
   ::||| :::: ||| ||| ||| ||| ||| ||| ||| |||
368 rglusargserleuvalthrgluserglylaargservalalapro 384
286 ACCAACAAGAAGCTAAATAARGAGCGGCTCCATTTAGCGCTGCTA 335
   :::: :::: :::: ||| ||| ||| ||| ||| ||| |||
385 leuaspysrtrpille...leuhsaparghlsleuuseralaser 400
336 TTGCTTCGTGAGAACATCGGAGCAAGACGCGCTGCGAGCGCATGT 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 lvalpheproleumethleuarg..... 408
386 ACAATACGCTGGAGAGTGTACAAAGTTCATCGAAGAGTCCATTCACG 435
   ::|| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
409 ..sergluleuuserlyrserasphellelvalleualathrprpe 424
436 CAGCG.....GACATCATCATATATCTGGCAGCTGCGCGGAGGT 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425 Thrllasprioglyasprhrllevaltrpmetlyleuasnvalglul 441
477 GCGCTACGACGATCGGAGCGGCTGCTTCGAGAGAGTGGCTGC 526
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 uasmetargargleuylsargglyarglyshlsgluserglyleua 458
527 CGCTTAAGTACCTTCAGAGCTG.....CATGAGTGGACGACG 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
458 spalaglytrleuylseryalasnspalalyrthslalavaltyrcys 474
565 GACGTGTGATACACGAGAGA...CGACCGAGTGTGACGAGTCTCATGT 611
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


```

158 trhisasprtmelnsenserglnphleglinsertleucluleuasp3 174
447 CATCATATATCTGGACACSTGGCGGAGTGCGGTACGAACGATCCGGC 436
|||||
174 yllellelyleuadarglathrprogluylscysleuasnarglleytl 191
497 AGCGGCGCTCTTCGAGAGAGCGCGCTGATACCTTCGAGAG 546
|||||
191 euargylargasnlglnlglnlelyleproleuclulyleuclulys 207
547 CTGCATGAGTTGCACAGACAGTGGTGCATACACAGAGACGCGGATC 596
|||||
208 leuistylrlyshisglusertprleucluleuistylrlyshis 224
597 GTGC.....AAGTCTAGTCTTCGATCCGATCTGAC 630
|||||
224 rphaesprtyrleuclnglnlvalprovalleuthrleuaspvalasn 239
seq_name: pir1:A38585

```

```

seq_documentation_block:
deoxycytidine kinase (EC 2.7.1.74) - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38585; S14321
R:Chotiner, E.G.; Shewach, D.S.; Datta, N.S.; Ashcraft, E.; Gibblin, D.; Ginsburg, D.;
Proc. Natl. Acad. Sci. U.S.A. 88, 1531-1535, 1991
A:Title: Cloning and expression of human deoxycytidine kinase cDNA.
A:Reference number: A38585; M0ID:91142207
A:Accession: A38585
A:Molecule type: mRNA
A:Residues: 1-260 <CHO>
A:Cross-references: GB:IM60527; NID:9181509; PIDN:AAA35752.1; PID:9181510
R:Eriksson, S.; Cederlund, E.; Bergman, T.; Joernvall, H.; Bohman, C.
FEBS Lett. 280, 363-366, 1991
A:Title: Characterization of human deoxycytidine kinase. Correlation with cDNA sequences
A:Reference number: S14321; M0ID:91192170
A:Accession: S14321
A:Molecule type: protein
A:Residues: 35-44, 'X', 46-47, 'X', 49-56; 163-168, 'X', 170-181; 223-236 <ERT>
A>Note: the amino end is blocked
C:Genetics:
A:Gene: GDB:DCX
A:Cross-references: GDB:126810; OMIM:125450
A:Map position: 4q13.3-q21.1
A:Superfamily: human deoxycytidine kinase
C:Keywords: ATP; blocked amino end; homodimer; phosphotransferase

```

```

alignment_scores:
Quality: 356.50 Length: 247
Ratio: 2.214 Gaps: 5
Percent Similarity: 65.182 Percent Identity: 30.769

```

```

alignment_block:
US-09-416-579A-1 x A38585 ..

```

```

Align seg 1/1 to: A38585 from: 1 to: 260

```

```

43 GCCGAGGACCCACCCCTTCACGCTCTCATCGAGGCGCAATCGGAG 92
|||||
16 SerGluglyThrArgIleLysLysIleSerIleGluglyAsnIleAla1 32
93 CGGGAAGACACGATTTGACCACTTCGAGAGTACAAAGACGACATT 142
|||||
32 actlylsserthrprhevalasnilleuclulyslneucysgluasprtpg 49
143 GCCTGCTGACGAGCCGCTCGAAGAGTGGCGCAACGTC..... 180
|||||
49 lvalvalprogluprovalalalargtrpcysasnvalglinserthrln 65
181 .....AACGGGTAAATCTGCT 197
|||||

```

```

66 asrglupheglugluleuthrmetsergluysasnlglyasnavalle 82
198 GGACGTGATGTACAAGATCCCAAGAGTGGGCGCATGCGCTTCGAGATT 247
|||||
82 uclimetyrlygluylsprogluargtrpserthrphelglnthr 99
248 ATGTCACGCTGACATGCTGACGTCGACACCGCCCAACAGCAAGAG 297
|||||
99 yralacysleuserlrglialarglalnleuAlaSerleuasnlglyls 115
298 CTAAA.....ATAGAGCGGCTCCATTTTAG 326
|||||
116 leuylaspralaglulysprovalleuprheprluargservaltyrse 132
327 CGCTGCTATGCTTCGTGAGACATCGAGCAAGAGCGCTGCGTGGAGC 376
|||||
132 rnsprtyrlyllephealaserasnleuylgluserglucysmelaag 149
377 AGGCGATGTACAATACGCTGAGAGAGTGGTACAGTTCATCGAGAGTCC 426
|||||
149 luthrclutprthrleuylrglnasprthrphiasprtmelnsasnlgln 165
427 ATT.....CACTGACGCGGACCTCATATATTCGCGACCTCGCC 470
|||||
166 phegluylinserleucluleuasp3lyllelyleuclulnaltthrpr 182
471 GGAGTGGCGGTACGAGACGCGTGGCGGCGCTGCTTCGAGAGAGCT 520
|||||
182 ogutprcysleuistargllyleuclulysnlglnlgln 199
521 GCGTGGCGCTTAAGTACCTTCACAGAGCTGCATGAGTTCGACGAGAC 570
|||||
199 lytleproleuclulyleuclulysleuclulysnlglnserthr 215
571 TPGATACACGAGAGGACGCG.....CACTGCGCAA 602
|||||
216 leuclenltharglthrleuclulysnprheasprtyrleuclulnvalpr 232
603 GGTCTAGTCTTCGATCGGATCGATCGAATCGAATTCGACGCGAGT 652
|||||
232 otleucluthrleuaspvalasn.....gluasprthelysasprlyst 246
653 ACCAGCGCTCGGAGACGACGATTTGACGCGCATCTGCAAGT 693
|||||
246 yrgluserleuvalgluylsvalylsgluprheuserthr 259
seq_name: pir2:JC6142

```

```

seq_documentation_block:
deoxyguanosine kinase (EC 2.7.1.113) precursor - human
C:Species: Homo sapiens (man)
C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 18-Jun-1999
C:Accession: JC6142
R:Johansson, M.; Karlsson, A.
Proc. Natl. Acad. Sci. U.S.A. 93, 7258-7262, 1996
A:Title: Cloning and expression of human deoxyguanosine kinase cDNA.
A:Reference number: JC6142; M0ID:96293511
A:Accession: JC6142
A:Molecule type: mRNA
A:Residues: 1-277 <JOH>
A:Cross-references: GB:U15668; NID:91477481; PIDN:AAC50624.1; PID:91477482
A>Note: It is uncertain whether Met-1 or Met-8 is the initiator
C:Comment: This enzyme is involved in mediating cytotoxicity of nucleoside analogs. I
Jnase, thymidine kinase 2.
C:Superfamily: human deoxycytidine kinase
C:Keywords: phosphotransferase
F:1-39/Domains: signal sequence #status predicted <SIG>
F:40-277/Product: deoxyguanosine kinase #status predicted <MAT>

```

```

alignment_scores:
Quality: 340.00 Length: 233
Ratio: 2.329 Gaps: 5

```


Percent Similarity: 62.661 Percent Identity: 33.047

alignment_block:

US-09-416-579A-1 x JG6142

Align seg 1/1 to: JG6142 from: 1 to: 277

```

49 GGCACCCAGCCCTTCACCGCTTCATCGAGGCGACATCCGCGACGGGAA 98
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 G1YArgG1ProArgArgLeuSerIleGluGlyAsnIleAlaValGlyLy 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 GACACGATTTGAACCTTCGAGATCAACAGCAACGATTCCTGCG 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 SSetThrPheValIleuLeuThrIleuProGluThrPheSerPhe 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 TGACCGACCGCGTCGAGAGTGGCGCAACGTCAAGGGGTA..... 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 LATHrGluProValAlaThrTrpGlnAsnIleGlnAlaAlaGlyAsnGln 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 .....AATCTGCTGAGCGGATGAGTACAA 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85 LysAlaCysThrAlaGlnSerLeuGlyAsnLeuLeuAspMetCysTrp 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 AGATCCCAAGAGTGGCGCATGCCCTTCAGAGTTATGTCACGTCGACA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 GGluproAlaArgTrpSerTrpThrPheGlnThrPheSerPheLeuSera 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 TGCTGCAGCTCGCACACCGCCCAACCAAGAGCTA..... 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 rGLeuValGlnLeuGluProPheProGluIleuLeuGlnAlaArg 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 .....AATTAARGAGCGCTCATTTTTCAGCGCTGCTATGCTT 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 LysProValGlnIlePheGlnArgSerValTrpSerAspArgTrpIleP 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 CGTGGAGAACATGCGACGAAAGCGCTGCTGAGCAGGCGATGTACATA 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 eAlaLysAsnLeuPheGlnAsnGlySerLeuSerAspIleGluTrpHis 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 CGCTGGAGAGTGTACAGTTTCATC.....GAAGAGTCCATTCACCTG 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
168 lEtyrGlnAspTrpHisSerPheLeuThrTrpGluPheAlaSerArgIle 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 CAGCGGAGCATCATATATATGCGACCTCGCGGAGGTGGCGTACGA 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 ThrLeuHisGlyPheIleTrpLeuGlnAlaSerProGlnValCysLeu 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 ACGCATCCGCGAGCGGCTGTTGTGAGAGAGCTGCGCGCTTAAGT 535
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 sArgLeuTrpGlnArgAlaArgGluGluGlyIleGluLeuAlaTr 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
536 ACCTTCAGAGAGCTGCATGATTCGACGAGACTGTTGATACACAGAGA 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 YrLeuGluGlnLeuHisGlyGlnHisGluAlaTrpLeuIleHisThr 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 CGACCG.....CAGTCGTGCAAGCTCTAGTCTTCGA 617
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
235 ThrIleuHisPheGlnAlaLeuMetAsnIleProValIleuValLeuAs 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
618 TGCCGATCTGAACCTGAAACATTTGCGACCGAGTACCGAGCTCGGAG 666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 pValAsn.....AspAspPheSerGluValThrIleGlnGln 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: p1r2:S71315
seq_documentation_block:
deoxyguanosine kinase (PC 2.7.1.113) precursor - human
C:Species: Homo sapiens (man)
C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: S71315; S78432
R:Mang, L.; Hellman, U.; Eriksson, S.
FEBS Lett. 390, 39-43, 1996
A:Title: Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA.

```

A:Reference number: S71315; MUID:96314545

A:Accession: S71315

A:Molecule type: mRNA

A:Residues: 1-260 <MAN>

A:Cross-references: EMBL:X97386; PIDN:CAA6054.1

A:Experimental source: tissue brain

R:Mang, L.; Hellman, U.; Eriksson, S.

submitted to the EMBL Data Library, April 1996

A:Description: Cloning and expression of human deoxyguanosine kinase cDNA.

A:Reference number: S78432

A:Accession: S78432

A:Molecule type: mRNA

A:Residues: 1-18, 'R', 20-260 <MAN>

A:Cross-references: EMBL:X97386; PIDN:CAA6054.1

A:Experimental source: brain

C:Genetics:

A:Gene: dgk

A:Genome: nuclear

C:Superfamily: human deoxycytidine kinase

C:Keywords: mitochondrial; phosphotransferase

F:1-32/Domain: transit peptide (mitochondrion) #status predicted <TRP>

F:23-260/Product: deoxyguanosine kinase #status predicted <MAT>

F:125-132/Region: DRS motif

F:185-191/Region: arginine-rich

alignment_scores:

Quality: 333.00 Length: 233

Ratio: 2.297 Gaps: 5

Percent Similarity: 62.232 Percent Identity: 32.618

alignment_block:

US-09-416-579A-1 x S71315

Align seg 1/1 to: S71315 from: 1 to: 260

```

49 GGCACCCAGCCCTTCACCGCTTCATCGAGGCGACATCCGCGACGGGAA 98
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18 G1YArgG1ProArgArgLeuSerIleGluGlyAsnIleAlaValGlyLy 34
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 GACACGATTTGAACCTTCGAGATCAACAGCAACGATTCCTGCG 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 SSetThrPheValIleuLeuThrIleuProGluThrPheSerPheLeuSera 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 TGACCGACCGCGTCGAGAGTGGCGCAACGTCAAGGGGTA..... 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 LATHrGluProValAlaThrTrpGlnAsnIleGlnAlaAlaGlyTrpGln 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 .....AATCTGCTGAGCGGATGAGTACAA 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 LysAlaCysThrAlaGlnSerLeuGlyAsnLeuLeuAspMetCysTrp 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 AGATCCCAAGAGTGGCGCATGCCCTTCAGAGTTATGTCACGTCGACA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 GGluproAlaArgTrpSerTrpThrPheGlnThrPheSerPheLeuSera 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 TGCTGCAGCTCGCACACCGCCCAACCAAGAGCTA..... 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 rGLeuValGlnLeuGluProPheProGluIleuLeuGlnAlaArg 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 .....AATTAARGAGCGCTCATTTTTCAGCGCTGCTATGCTT 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 LysProValGlnIlePheGlnArgSerValTrpSerAspArgTrpIleP 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 CGTGGAGAACATGCGACGAAAGCGCTGCTGAGCAGGCGATGTACATA 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 eAlaLysAsnLeuPheGlnAsnGlySerLeuSerAspIleGluTrpHis 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 CGCTGGAGAGTGTACAGTTTCATC.....GAAGAGTCCATTCACCTG 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 lEtyrGlnAspTrpHisSerPheLeuThrTrpGluPheAlaSerArgIle 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 CAGCGGAGCATCATATATGTCGCGACCTCGCGGAGGTGGCGTACGA 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

seq_name: pir2:E96743

seq_documentation_block:

probable deoxyguanosine kinase F28P5.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:\date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:\Accession: E96743

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.

ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408: 815-820, 2000

A: Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
Issue: 408, 610-620, 2000

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A863441: MIMD:21016719

A; Accession: E96743

A;status: preliminary
A;molecule type: DNA

A:Residues: 1 361 <STO>
A:Cross-references: GB:AE005173; NID:q11054411; PIDN:AAG27798.1; GSPDB:GN001411

C;Genetics:
A;Gene: F28P5.7

A;Map position: 1

alignment_scores:

Quality:	315.50
Ratio:	1.960

Percent Similarity: 57.706 Percent Identity: 29.032

alignment_block:
ITS-09-416-579A-1 v F067A3

Align seg 1/1 to: E96743 from: 1 to: 361

49 GGCACCCAGCCC.....TTCACCGTCTCATCGA 77

```

||| :::|||
36 G!vProG!uProAsn!euAsnVal!velvswAr!authr!bhc!usVal!C! 53

```

```
78 GGGGAAAGATCCCCCAACCGAAGCCGAATTGC    206
```

```

10 GGGGACATCGGACGGGAGGACCACGTAATGG.....AACG 11
    |||||::: |||||:::||||: |||:

```

52 uGlyAsnIleSerValGlyGlySerThrPheLeuGlnArgIlealaasng 69

116 ACCTTGAGAGTACAAGAAGACATTTGCCCTGCTGACCGAGCCCGTGAG 16

```
69 lUthrvAlGlULeGlInAsplEuVaGlUleVaPlroGlUProVaAlAsp 85
```

166 AAGTGGCGCACGTC.....AACGGGTAATCTGCTGGAGCTGATGTA 20

86 LysTrpGlnaspValGlyProAspHisPheasnIleuAspAlaphety 10

THE
JOURNAL
OF
THE
ROYAL
ANTHROPOLOGICAL
INSTITUTE
OF GREAT
BRITAIN
AND IRELAND
PART 1
1906

٧٠

```

210 CAAGATCCCAAGAAAGTGGGCATGCCCTTCAGAGTATGATCAGCTGA 259
102 TSGEupProGlnIupArgIupAlaIupThrIupPheGlnAsnIupValIupPheVal 119
260 CC.....ATGCTCAGTGGCACACCGCCCAACCAACCAAGAAAGCTAA 303
119 hrArgIupMetGlnIupGlnIupSerIupAlaIupSerGlnValIupProIupArg 135
304 ATAAAGACGGCTCCATTTTACGCTGCTATTCCTTCCTGGAGAACAT 353
136 LeuMetGlnIupSerValIupSerAspArgMetValIupPheValIupArgAlaVal 152
354 GCGACGAACGGCTCGCTGGAGCAGGGGCATGTACAAATACCTGGAGAGT 403
152 IHisGlnAlaIupTrpMetAsnIupMetGlnIupLeuIupSerIupLeuIupAspSer 169
404 GGTACAGATTTCATCGAAGAGTCCATT...CACCTGCAGCGGCAGACCTCATC 450
169 rPheAspProValIupSerIupLeuProGlnIupLeuValIupProAspGlnPhe 185
451 ATATATCTGGCAGCTGCCGAGAGTGGCTACGAAACGATCCGGACGC 500
186 ILeuTrpLeuArgIupAlaSerProAspTrpIupCysIupAsnIupMetIupLeu 202
501 GGCCTGCTTGGAGAGAGCTGCGTGGCGCTTAACTACCTCAGAGACTGC 550
202 GlysArgAlaGlnIupGlnIupGlnIupAlaIupSerIupLeuIupGlnAsnIupLeu 219
551 ATGAGTTGCACACGACTGGTTG..... 573
219 IAsnIupLysIupHisIupSerIupTrpIupLeuIupProIupGlnIupSerGlnAsnIupGln 235
574 ...ATACACACAGACGACCG..... 591
236 ValIupSerValIupSerIupProIupSerIupLeuIupHisIupMetAspIupSerIupHisIup 252
591 ..... 591
252 oAspIupIleIupAspArgValIupPheIupTrpIupLeuIupGlnIupGlnAsnIupMetIupHisIup 269
592 ....CAGTGGTCAAGGCTCTAGTCTCGATGCCGATCCGATCACTGGAGAA 636
269 erIupGlnIupValIupProIupAlaIupLeuValIupLeuAspIupGlnIupProIupAsnIupLeuAsp 285
637 .....AACATTGGCACCGAGTACACACGCTCGGAGCGACG 671
286 PheSerIupArgIupIleIupGlnAlaIupLysIupThrGlnIupValIupArgIupGlnIupValIupAlaIup 302
672 CATATTGCAGCGCATCTCAAGTAAACCAACAGCGCTCG 708
302 upPheIupGlnIupPheValIupLysIupGlnIupIupTrpSer 314
seq_name: p1r2.D35216

```

seq_documentation_block:

FP26 protein - fowlpox virus (strain FP-1)
C;Species: fowlpox virus

C;Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 18-Jun-1999
C;Accession: D35216

R; Tartaglia, J.; Winslow, J.; Goebel, S.; Johnson, G.P.; Taylor, J.; Paoletti, E.
J. Gen. Virol. 71: 1417-1424 1990

J. Gen. Virol. 71, 1517-1524, 1990

A;Reference number: A35216; MUID:90324937
A;Accession: D35216

A;Status: preliminary
A:Molecule type: DNA

A;Residues: 1-219 <TAR>

Cross references: GR:XR1202; NID:g61222; PIDN:CAA35067.1; PID:g61227
C/Superfamily: human deoxycytidine kinase

alignment_scores: 244.50
Quality: 244.50

210
Lenglin: 244:30
Xinsheng: 244:30

Ratio: 1.940 Gaps: 10
Percent Similarity: 57.534 Percent Identity: 31.507
US-09-416-579a-1 x D35216 ..

Align seg 1/1 to: D35216 from: 1 to: 219

```

73 ATCGAGGCGGACATCGGCGGCGGAGACGATATTGAACTTGA 122
|||||
14 IIEGLIGLYASnIleSerSerIlyThrAspValIleuAsnIleuAr 30
123 GAAGTACAAAGAAC.....GACATTGCGCTGACCGAGC 157
|||||
30 gAsnIleAsnValIalSerPheHISAspValGluAspArgTyrThr 47
158 CCGTGAGAGAGTGGCGCAACGTCAGAGGGGTAAATCTGTGAGCTGATG 207
|||||
47 rOIleGIuIyS.....GluLeuIleArgLysPhe 56
208 TACAAGATCCCAAGAGAGTGGCGCATGCCCTTTCAGAGTATGTCACGCT 257
|||||
57 HIsGIuAsnProSerArgTyrPserIyAlaLeuGlnThrHISYrCysMe 73
258 GACCATGCTGCAGTGCACACC.....GCCCAACCAACAAGAACG 298
|||||
73 LysArgValArgMetHISLeuGlyCysPheValProSer.....ArgV 88
299 TAAATATATAGAGAGCGCTCATTTTACGGCTATGCTATGCTGTGAG 348
|||||
88 aIAsnIleLeuGlnArgSerIlePheSerAspArgTyrValPheAlaGlu 104
349 AACATGCGACGAACCGCTGCTG.....GAGCAGGCGCATGTACAA 389
|||||
105 AlalArThrAlaLeuGlyTyrMetAspAspProGluThrAlaLeuTyrCy 121
390 TACGCTGAGAGAGTGTACAGATCATCGAGAGAGTCCATTACCTGACG 439
|||||
121 sLySgInHISAspTyrPyr.....ThrAspLysLeuGlnIleGlnP 135
440 CGGACCTCATATATCTGCGCACCTGCGGAGAGTGGCGGTACGACGC 489
|||||
135 heAspGlyIleIleTyrLeuArgThrIleProGluSerCysLysGluArg 151
490 ATCCGGCAGCGGCTGCTGTGAGAGAGCTGC.....GTGCCGCTTAA 533
|||||
152 ILeAsnGluLysSerIleThrGluLysAsnTyrProAsnIleSerIleAs 168
534 GTACCTCAGAGAGTGCATGAGTGCACGACGAGAGTGTGATACACAGA 583
|||||
168 pTyrlLeuLysThrLeuHISGluLysHISGluLeuTyrPleu..... 181
584 GACGACCGCAGCTGTCGACG.....GTCTAGTCTCTGATGCCGAT 624
|||||
182 .....ThrGlnCysLysLysValProValIleuIleIleAspGlyGlu 195
625 CTGAACTGGAACAACTGGCAGCCGAGTACACGCGCTCGGAGAGACAT 674
|||||
196 .....GluAspPheIle 199
675 ATTGCAC 681
|||||
199 ePheAsp 201

```

seq_name: plr1:KIBETC

seq_documentation_block:

thymidine kinase (EC 2.7.1.21) - Ictalurid herpesvirus 1

C:Species: Ictalurid herpesvirus 1

A:Note: host Ictalurus punctatus (channel catfish)

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 11-Jun-1999

C:Accession: J01336

R:Harrison, P.T.; Thompson, R.; Davison, A.J.

J. Gen. Virol. 72, 2583-2586, 1991
A:Title: Evolution of herpesvirus thymidine kinases from cellular deoxycytidine kinas

A:Reference number: J01336; M01D:92013982

A:Accession: J01336

A:Molecule type: DNA

A:Residues: 1-228 <HAR>

A:Cross-references: GB:M75136; NID:9331209; PIDN:AAA8108.1; PID:9331215

C:Superfamily: Ictalurid herpesvirus thymidine kinase

C:Keywords: ATP; DNA biosynthesis; phosphotransferase

alignment_scores:

Quality:	Ratio:	Length:	Gaps:
212.50	2.063	194	7
Percent Similarity: 53.093	Percent Identity: 34.021		

alignment_block:

US-09-416-579a-1 x KIBETC ..

Align seg 1/1 to: KIBETC from: 1 to: 228

```

46 GAGGCGACCGAGCCCTTACACGCTCTC.....AT 74
|||||
5 GluGlyProThrProAlaArgPheLeuProGluGlyLeuValPheCysVa 21
75 CGAGGCGAATCGGCGGCGGAGACGATTTTGAACCTTGAGAGA 124
|||||
21 IGluGlyAsnIleGlyCysGlyLysSerThrLeuValLysAlaLeuMeG 38
125 AG.....TACAAGAACGATTTGCTGCTGAGCAGCGCGCGAGAG 168
|||||
38 LuArgValAlaGlySerGlyValAsnValValGluLeuProValAspGln 54
169 TGCGGCACCTCAACGGGTAATCTGCTGAGCTGATGATACAAAGATCC 218
|||||
55 TrpValAsnHISAsnGlyLysAsnTyrLeuGluLeuSerTyrThrAspR 71
219 CAAGAAGTGGCCATGCCCTTTCAG.....AGTATGTCA 253
|||||
71 oThrGlyTyrAlaValProPheGlnAsnLeuValPheAspSerTyrValAla 88
254 CGCTGACATGCTGAGTGCACACCGCCCAACCAAGAGTAATAA 303
|||||
88 snValGlnArgLeuGlnAsnPro.....Asp 96
304 ATATAGAGCGCTCCATTTTACGCTGCC.....TATGCTTGTGGA 347
|||||
97 ILeMeGluArgSerProMetSerAlaThrArgValPheCysAlaValAs 113
348 GAACATGCGACGA.....AACGCTCGCTGAGCAGGCGCATGTACAAATA 391
|||||
113 nGlySerArgGlyValIleProAlaThrAlaLeuProGlyMetAlaValAla 130
392 CGCTGAGAGAGTGTACAAATTCATCGAAGAGTCCATTACCTGACGCG 441
|||||
130 rGlyGluAlaValMetArgThrIle.....AlaThrArg 141
442 GACCTCATATATCTGCGCACCTGCGGAGGTGGCTTACGAACGAT 491
|||||
142 ProValPheValTyrLeuGluLeuProProGluGluCysLeuArgArgMe 158
492 CGGCGACGCGGCTGCTTCTGAGAGAGAGTGCCTGCGCGCTTACGATCTG 541
|||||
158 LArgArgArgAspArgThrGlyGluAlaGlyValGlyLeuAspTyrLeuA 175
542 AGGAGCTGATGATTGACACGAGGAGCTGTTG 573
|||||
175 rGluLeuHISGluArgTyrGluAlaTyrPleu 185

```

seq_name: plr2:T03086

seq_documentation_block:

probable thymidine kinase (EC 2.7.1.21) - Chilo iridescent virus


```

149 sTyrtyrSer.....llepheasnlystrhrasnserrhrproi 163
      :          :||| :          :|||
      :          :||| :          :|||
605 TCCTAGTCGATCGACGATCGACGACGAAACATATGCGACCGACGATAC 654
      :          :||| :          :|||
163 lclYpHelelSpalaserlysaasnlgualvalgluclualsphe 179
      :          :||| :          :|||
seq_name: p1r1:T42934

```

seq_documentation_block:

```

thymidine kinase (EC 2.7.1.21) [similarity] - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T42934
R:Albrecht, J.C.; Fleckenstein, B.
  submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z22274
A:Accession: T42934
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-527 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AC95545.1
A:Experimental source: strain 73
C:Genetics:
A:Note: ori 20
C:Superfamily: salmistrine herpesvirus 1 thymidine kinase; herpesvirus thymidine kinase
C:Keywords: phosphotransferase

```

alignment_scores:

```

Quality: 149.50 Length: 254
Ratio: 1.133 Gaps: 10
Percent Similarity: 51.969 Percent Identity: 25.197

```

alignment_block:

US-09-416-579A-1 x T42934 ..

-Align seg 1/1 to: T42934 from: 1 to: 527

```

40 TACGCGGAGGCGACCCACCTTCACCGCTCTCATCGAGGCGACATCGG 89
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
203 HisGngluProlleProAlaPhePheleuGluGlySerIleGI 219
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
90 CAGCGGGAAGACACGATATTGAACACACTTCGAGAGATAC.....A 130
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
219 yValGlyythrhrleuLeuLysSerMetArgGlyIleleuProGlyL 236
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
131 AGAACGACATTTGCTCTCTGACCGACCGCTCGAAGATGGCGGACATC 180
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
236 ysaAn...ValleuThrPheHISgluPromeAlaPheTrpLysAsnVal 251
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
181 AACGGGTAATTCGTGAGACTATGTACAAA.....GATCCCAAGAA 224
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
252 PheSer...AsnSerLeuAspGluValTyrLysLeuThrleuProAla 267
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
225 GTGGGCGATCGCTTTCAGAGTTATGTACGCTGACCATGCTGACGTCG 274
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
267 sValGlySerMetThrAsnSerLysLeuLeuAlaCysGlnleuLysP 284
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
275 ACACCGGCCCA.....ACCAACAAGAAAGCTAAATA 306
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
284 heAlaAlaProlleuLeuAlaLeuLysThrSerThrAspArgLeuSer 300
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
307 ARGAGGCGCTCATTTTACGCGCTGCTATTGCTTGTGAGAGAAACATCG 356
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
301 SerAsnArgSerAsnLeuSerSerAsnMetTrpValMetPheAspArg 317
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
357 ACGAACGCGCTCGCTGAGACGAGGACATGACAAATACGCTGAGAGAGT 406
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
317 sProlleuSerAlaThrValValPheProTyrMetHisTyrGlnAsncl 334
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

407 ACAAGTTCATCGAGAGTCATTCACCTG..... 435
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
334 heLeuSerPheSerHisLeuValHisLeuTrpLysPheArgAlaSer 350
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
436 CAGCGGACCTCATATATATCTGCGACCTCGCGAGGCTGCGATACGA 485
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
351 HisGlyAspAsnIleIleleuLeuAsnLeuAsnSerGlnGlnleuGI 367
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
486 ACGCATCCGCGACGCGCTGTTCTGAGAGACCTGCGCTCCCTTAAGT 535
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
367 uArgValLysArgArgAsnArgLysgluGluLysCysValSerleuGlu 384
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
536 ACCTTCAGAGCTG.....CATGATTGCACCGACGCTGTTG 573
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
384 iStleArgMetLeuAsnSerCysTyrHisAlaValTyrCysAlaTrpLeu 400
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
574 ATACACCGACAGA...CGACCGAG..... 594
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
401 leuValGlnAsnPhetrhrProGluGluIleValGluValCysPheAsn 417
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
595 .....TCGTGCAGAGTCTAGTCTCGATCGCG 622
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
417 rLysGlnIleThrAspLeuArgSerSerLysProserPheleuGluLys 434
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
623 ATCTGACCTGAAACATATGCGACGAGATACGAGGCTCGAGACGACG 672
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
434 lSvalSerIleGluAspVal.....LeuLysSerSer 444
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
673 ATATTGCGACGCC 684
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
445 llePheAsnAla 448

```

seq_name: p1r2:F75535

seq_documentation_block:

```

deoxyguanosine kinase/deoxyadenosine kinase subunit - Deinococcus radiodurans (strain
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75535
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
  S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
  Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75535
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <WHI>
A:Cross-references: GB:AF001891; GB:AF000513; NID:96457973; PIDN:AAF09882.1; PID:9645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0298
A:Map position: 1
C:Superfamily: Lactobacillus acidophilus deoxyadenosine kinase

```

alignment_scores:

```

Quality: 148.00 Length: 188
Ratio: 1.345 Gaps: 6
Percent Similarity: 58.511 Percent Identity: 26.064

```

alignment_block:

US-09-416-579A-1 x F75535 ..

-Align seg 1/1 to: F75535 from: 1 to: 207

```

67 GTCCTCATCGAGGCGACATCGCGACGCGAGACGACGATATTGACCA 116
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3 ValAlaValSerClyAsnIleLysSerGlyLysSerThrLeuThrArgMe 19
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
117 CTTC...GAGAGTACACAGACGACATTTGCTGCTGACGAGCCGCTCG 163
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

19 tleuAlaGluArgTyrGly.....LeuArgProValT 30
164 AGAAGTGGCGCAAGCTCAACGGGGTAAATCGCTGAGCTATGTACAAA 213
   ::::: ||| ||| ||| ||| :::::
30 yrgLupProGlyAlaGluasn.....ProGlyLeuGluAspPheTyrHis 44
214 GATCCCAAGAAAGTGGGCGCATGCCCTTTCAGAGTATGTACAGCTGACCAT 263
   ||| ::::: ||| ||| ||| |||
45 AspMetAlaGlnTyrSerPheHisSerGlnValTyrPheLeuSerArgAr 61
264 GCTGCAAGTGCACACCGCC...CCAACCAACAAGAAGCTAAATAAARGG 310
   ||| ::::: ||| ||| ||| |||
61 gLeuGlnGlnHisLeuGlyMetValThrGlyAlaArgTyrValIleGln 78
311 AGCGCTCCATTTTAAAGCTGCTATGCTTGTGTGGAGAACATGGCAGCA 360
   ::| ||| ::::: ||| ||| ||| |||
78 sPArgThrValPheGluAspAlaAsnIlePheAlaArgAsnLeuTyrGlu 94
361 AAGCGCTGCTGAGCAGCGCATGTACAAATACGCTGAGAGAGTGGTACAA 410
   ::| ||| ::::: ||| ||| ||| |||
95 SerGlyGlnMetGlyGlnArgAspTrpGlnThrTyrArgGlyLeuTyrCl 111
411 GTTCATCGAAGATCCATTACCTGACGGGAGCTCATATATATCTGC 460
   : ::::: ||| ||| ||| ||| ||| |||
111 uGlyValLeuProAlaLeuArgVal...ProAspLeuIleHisIleA 127
127 sPAlaGlyLeuProThrLeuArgArgArgIleAlaLeuArgGlyArgasp 143
511 GAGGAGAGCTGCGTGGCTTAAAGTTCATGAGAGCTCATATATATCTGC 560
   ||| ::::: ||| ||| ||| ||| ||| |||
144 TyrGlnGlnAlaIleProAspGluTyrLeuAlaGlyLeuAsnArgLeuTyr 160
561 CCAGAGCTGGTGAATACACCAAGAGACCGAGCTGTCGAAGTCTCAG 610
   : ||| ::::: ||| ||| ||| ||| |||
160 rAlaGlyTyrPile.....AlaAlaPheAspLeuGlyCysProValValA 174
611 TCCTGATGCCGAT 624
174 rGValAspGlyasp 178
seq_name: p1r2:F86769

```

```

seq_documentation_block:
deoxyribose kinase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: F86769
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. In press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: F86769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <STO>
A:Cross-references: GB:AE005176; NID:q12724122; PIDM:AAK05256.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: dukB
C:Superfamily: Lactobacillus acidophilus deoxyadenosine kinase

```

```

alignment_scores:
Quality: 141.00 Length: 181
Ratio: 1.382 Gaps: 8
Percent Similarity: 56.354 Percent Identity: 27.624

```

```

alignment_block:
US-09-416-579a-1 x F86769 ..
Align seg 1/1 to: F86769 from: 1 to: 212

```

```

67 GTCTCATGAGGCAACATGCGAGCGGAGACCAAGCATTTGACCA 116
   ::::: ||| ||| ||| ||| ||| ||| |||
2 IleValIleuAlaGlyThrIleGlyAlaGlyLysSerSerLeuAlaLysAl 18
117 CTTC...GAGAAGTACAAAGACATTTGCTGCTGACCGAGCCGCTGC 163
   ::| ||| ||| ||| ||| ||| |||
18 aLeuGlyGlnHisLeuTyrThrAspAlaPhe..... 28
164 AGAAGTGGCGCAAGCTCAACGGGGTAAATCGCTGAGAGCTGATGACAA 213
   ::::: ||| ||| ||| ||| ||| |||
29 ....TyrGlnAlaValAspAsnAsnProValLeuAspLeuTyrGln 43
214 GATCCCAAGAAGTGGGCGCATGCCCTTTCAGAGTATGTACAGCTGACCAT 263
   ||| ||| ||| ||| ||| ||| |||
44 AspProGlnTyrTyrAlaPheLeuLeuGlnIleTyrPheLeuAsnLysAr 60
264 GCTGCAAGTGCACACCGCCCAACCAACAAGAAGCTAAATAAARGGAGC 313
   ::::: ||| ||| ||| ||| ||| |||
60 gPheGlnSerIleLysMetAlaTyrArgGlnAspAsnValLeuAspA 77
314 GCTCATTTTAAAGCTGCTGCTATGCTTGTGGAGAACATGGCAGCAAC 363
   ||| ||| ||| ||| ||| ||| |||
77 rGSerIlePheGlnAspGluLeuPheLeuThrLeuAsnTyr...LysAsn 92
364 GGCTGCGTG.....GAGCAGGCGCATGTAC.....AA 389
   ||| ::::: ||| ||| ||| ||| |||
93 GlnAsnValThrLysThrGlnLeuGlnIleTyrGlnAspLeuAlaAs 109
390 TAGCTGAGAGAGTGTACAAATTCATGAGAGCTCATTCACCTGCAGG 439
   ||| ||| ||| ||| ||| ||| |||
109 mLeuLeuGlnLuu.....LeuAspGlyMetProLysLysArgp 122
440 CGGACCTCATCATATATCTGCGCACCTGCGCGAGGTGGCGTACGAGCG 489
   ||| ||| ||| ||| ||| ||| |||
122 roAspLeuLeuValTyrIleAspValSerPheGlnTyrMetLeuSerArg 138
139 IleAlaGlnArgGlyArgSerPheGlnGlnIleGlnAsnGlnGlnLys 155
531 TAAG...TACCTTCAAGAGCTGATGAGTTCACACGAGACTGG 570
   ||| ||| ||| ||| ||| ||| |||
155 uLysAspTyrTyrAlaGlnValHisAspGluTyrProAspTrp 169
seq_name: p1r1:K18BHS

```

```

seq_documentation_block:
thymidine kinase (EC 2.7.1.21) - salmireline herpesvirus 1 (strain 11(Onc))
C:Species: salmireline herpesvirus 1
A:Note: host Salmireline scituens (common squirrel monkey)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 19-Jan-2001
C:Accession: A33374
R:Hones, R.W.; Craxton, M.A.; Williams, L.; Gompels, U.A.
J. Gen. Virol. 70, 3003-3013, 1989
A:Title: A comparative analysis of the sequence of the thymidine kinase gene of a gam
A:Reference number: A33374; M01D:90063548
A:Accession: A33374
A:Molecule type: DNA
A:Residues: 1-527 <HON>
A:Cross-references: EMBL:DO0543; NID:g221853; PIDN:BAH00432.1; PID:g221856
C:Superfamily: salmireline herpesvirus 1 thymidine kinase; herpesvirus thymidine kinas
C:Keywords: ATP; DNA biosynthesis; nucleotide binding; P-loop; phosphotransferase
F:209-492/Domain: herpesvirus thymidine kinase homology <HK>
F:216-223/Region: nucleotide-binding motif A (P-loop)
F:311-315/Region: nucleotide-binding motif B
F:222/Binding site: ATP (Lys) #status predicted

```

```

alignment_scores:
Quality: 134.50 Length: 257
Ratio: 1.027 Gaps: 13
Percent Similarity: 50.973 Percent Identity: 25.292

```

us-09-416-579a-1.rpr

US-09-416-579A-1 x KIBEHHS

deoxyguanosine kinase (EC 2.7.1.113) - Lactobacillus acidophilus
N:Alternate names: dGua kinase subunit
C:Species: Lactobacillus acidophilus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 18-Jun-1999
C:Accession: D51610.D54357

```

deoxyguanosine kinase (EC 2.7.1.113) - Lactobacillus acidophilus
N:Alternate names: dgno kinase subunit
C:Species: Lactobacillus acidophilus
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 18-Jun-1999
C:Accession: B56168; D54257
J:Ma, G.T.; Hong, Y.S.; Ives, D.H.
J. Biol. Chem. 270, 6595-6601, 1995
A>Title: Cloning and expression of the heterodimeric deoxyguanosine kinase/deoxydenc
A:Reference number: A56168; MUID:9520449
A:Accession: B56168
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <MAA>
A:Cross-references: GB:U01881; NID:g403982; PIDN:AAB09751.1; PID:g403984
A:Experimental source: Strain R-26
R:Iveda, S.; Ma, G.T.; Ives, D.H.
Biochemistry 33, 5328-5334, 1994
A>Title: Heterodimeric deoxynucleoside kinases of Lactobacillus acidophilus R-26: fun
A:Reference number: A54257; MUID:94227067
A:Accession: D54257
A>Status: preliminary
A:Molecule type: protein
A:Residues: 2-18 <IKK>
A:Experimental source: R-26
A>Note: sequence extracted from NCBI backbone (NCBIP:146751)
C:Superfamily: Lactobacillus acidophilus deoxydensline kinase
C:Keywords: phosphotransferase

alignment_scores:
    Quality: 134.00      Length: 212
    Ratio: 1.039        Gaps: 8
    Percent Similarity: 60.849   Percent Identity: 22.642

alignment_block:
US-09-416-579A-1 x B56168 ..

Align seg 1/1 to: B56168 from: 1 to: 224

67 GTCCATCATCGAGGCGACATCGCGAGCGGAAGAACCATATTGAACCA 116
   ::::::::::::::||| ||||||:::||||:~::~
4  llevalleuserglyproilleglyalglysserSerleuthnGlyI1 20
   :::::||||| | ||| ||::|||
20 eleuserltyr.....LeuglythrAsnProphe... 30
   :::::||||| | ||| ||::|||
167 AGTGGCCACGCCTCACGCCGGGTAAATCTGCTGAGAGCTGATACAAGAT 216
   .....|||::: ||||| ||||||:~:::::
31 ..TyrgluserValAspaSpasnpProvalleuProleuphetylgluasn 46
   .:::|||||:~:::::
217 CCCAAGAAGTGGCCATGCCCTTCAGAGTTANGTCACGCTGACCATGCT 266
   ||||||||:~::::: ||||| ||| :
47 ProlyslystyrAlaPheLeuLeuglnValtyrPheLeuasnThrArgph 63
   ::::||| :~:::::
267 GCAGTCCACACCGCCCCAACACACAGAAGAAGCTAAATAARGAGCGCT 316
   ::::||| :~:::::
63 enryserlleysSerAlaleuthnraspasnAsnvalleuasparys 80
   eaygSerlleysSerAlaleuthnraspasnAsnvalleuasparys 80
317 CCATTTTAACCGCTCGCTATTGCTTCGTGGAGAAACATGACGACAAGCGC 366
   |||||:~::::: ||| :
80 eriletyrgluaspalaLeuphe.....GlmelAsnAla 92
   ::::||| :~:::::
367 TCGGTGAGAGCGGCAATGATACATATACGCTGAGAGAGTGATACAATTGAT 416
   :::::~:::::
93 AspllelYlrfgAlatrProgluglnValAspThrtYtyrGluleue 109
   :::::~:::::
417 CGAAGAGTCCATT.....CACCTGACAGCGGAGC 445
   :::::~:::::
109 uhtASnmelmetsercLluenuasparymetProplyslsAsnPromspl 126
   :::::~:::::
446 TCATCATATATCTGCGACACTCGCGGAGGTGGCGTAGCAAAACCATCCGG 495

```

126 euleuValHisIleaspValserTyrAspThrMetLeuLysArgIleGln 142
 496 CAGGGGGCTGCTGAGGAGAGTCGCCGCTTAGTACCTTCAGGA 545
 143 LysArgIleGlnAsnTyrGlnGln.....LeuSerTyrAspProth 156
 546 GCTGCATGAGTTGCACGAGACTGTGTATACACAGACAGACCGCAGT 595
 156 rleuGluAspTyrTyrLysArg...LeuLeuArgTyrTyrLysProThr 172
 596 CGTCCAGGTC.....CTAGCCTCGATGCCGATCTGAC 630
 172 yralAlaLysTyrAspTyrSerProLysMetThrIleAspGlyAsp...Lys 187
 631 CTGGAAAAATTTGGCACCAGTACCGCTCGGAG 666
 188 LeuAspMetAlaSerGluGluAspArgGlnGlu 199

seq_name: p1r1:G70129

seq_documentation_block:

deoxyguanosine/deoxyadenosine kinase I (EC 2.7.1.-) subunit 2 dck homolog - Lyme disease
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: G70129
 R/Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 R:Authors: Smith, H.O.; Venter, J.C.
 A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A/Reference number: A70100; MID:98065943
 A/Accession: G70129
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 R:Molecule type: DNA
 A/Residues: 1-205 <KLE>
 A/Cross-references: GB:AE001134; GB:AE000783; NID:q2688133; PIDN:AAC66630.1; PID:q268813
 A/Experimental source: strain B31
 C:Superfamily: Lactobacillus acidophilus deoxyadenosine kinase
 C/Keywords: phosphotransferase

alignment_scores:
 Quality: 132.50 Length: 196
 Ratio: 1.380 Gaps: 9
 Percent Similarity: 48.980 Percent Identity: 24.490

alignment_block:
 US-09-416-579A-1 x G70129 ..

Align seg 1/1 to: G70129 from: 1 to: 205

67 GTGCTCATGAGGCAACATCGGCGAGCGGGAAGACACAGTATTGAC.. 114

 7 IleValIleGluGlyLeuIleGlyValGlyLysThrThrLeuGlyAsnI 23
 115CACTTCGAGAGTACACAGAGACGACA 139
 23 eLeuSerLysGluLeuGluValProPheTyrSerGluLeuAsnAsnAsp 40
 140 TTTGCTGCTGACCGAGCCGTCGAGAGAGTGCGCAACGTCACAGGGGTA 189
 ..
 40 heThrLeuAla..... 43
 190 AATCTCTGAGAGTGATGTACAAGATCCCAAGAGTGGCCATGCCCTT 239

 44 ...ValLeuAspLysPheTyrLysAspLysSerArgTyrPalaPheProVa 59
 240 TCAGAGTTATGTACGCTGACATGCTGCAGTCGACACCGCCCAACCA 289

 59 IeGln.....IleAsnPheLeu.....A 65
 290 ACAAGAGCTAAATAATATATAGGAGCGCTCCATTTTAGCGCTGCTATGCG 339

|||||.....
 65 snGluArgPheLysLeuIle...LysGlyValPheArgThr..... 77
 340 TTCGTGAGACATCGACGAACCGCTCGAGACGAGGCATGTAC.. 387
 78LysGlyLysIleLeuAspArgSerIleTyrG1 88
 388AATACGCTGAGAG 400
 88 yAspCysValPheAlaSerLeuLeuAsnCysAspGlyHisIleSerAspG 105
 401 AGTGATACAGTTTCATCGAAGATCCATT.....CACTTG 435
 105 IuGluTyrLysIleTyrIleAspLeuLeuAspAsnMetLeuGluHisSer 121
 436 CAG...GGGACCTCATCATATATCTGCCACCTGCGGAGGTGGGTA 482
 122 GlnArgProSerLeuLeuValTyrLeuAspCysSerIleAspGluValG1 138
 483 CGAAGCATCCGGCAGCGGGCTGTTCTGAGGAGAGCTGCTGCCGCTTA 532
 138 uArgArgIleLysAsnArgAsnArgSerPheGluMetAsnIleProArgA 155
 533 AGTACCTCAGAGACTGCATGAGTTGCACACGAGACTGG 570
 155 spTyrLeuGluGlyLeuAsnArgLysTyrLeuLysTrp 167

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 05:42:59 ; Search time 48.33 Seconds

(Without alignments)
2949.540 Million cell updates/sec

Title: US-09-416-579a-1
Perfect score: 753
Sequence: 1 atgagcgagagcagcatctg.....gccagaggtccagataa 753

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 segs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCUTS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46.8	6.2	9551	1 US-08-056-200-93	Sequence 93, Appl
2	46.8	6.2	9551	2 US-08-800-644-93	Sequence 93, Appl
3	41.2	5.5	36519	3 US-08-923-137-2	Sequence 2, Appl
4	39.2	5.2	467	2 US-08-476-176B-17	Sequence 17, Appl
5	39.2	5.2	467	3 US-08-127-721A-17	Sequence 17, Appl
6	39.2	5.2	467	3 US-08-485-246A-17	Sequence 17, Appl
7	39.2	5.2	468	2 US-08-476-176B-13	Sequence 13, Appl
8	39.2	5.2	468	3 US-08-127-721A-13	Sequence 13, Appl
9	39.2	5.2	468	3 US-08-485-246A-13	Sequence 13, Appl
10	38.2	5.1	467	2 US-08-476-176B-15	Sequence 15, Appl
11	38.2	5.1	467	3 US-08-127-721A-15	Sequence 15, Appl
12	38.2	5.1	467	3 US-08-485-246A-15	Sequence 15, Appl
13	38.2	5.1	468	2 US-08-476-176B-11	Sequence 11, Appl
14	38.2	5.1	468	3 US-08-127-721A-11	Sequence 11, Appl
15	38.2	5.1	468	3 US-08-485-246A-11	Sequence 11, Appl
16	38.2	5.1	7218	1 US-08-232-463-14	Sequence 14, Appl
17	37	4.9	44377	2 US-08-804-227C-7	Sequence 7, Appl
18	37	4.9	44377	2 US-08-804-198-1	Sequence 1, Appl
19	36.4	4.8	831	1 US-07-882-329-1	Sequence 1, Appl
20	36.4	4.8	831	1 US-08-183-213-1	Sequence 1, Appl
21	36.4	4.8	1120	1 US-07-882-329-3	Sequence 3, Appl
22	36.4	4.8	1120	1 US-08-183-213-3	Sequence 3, Appl
23	36.4	4.8	1329	1 US-07-882-329-4	Sequence 4, Appl
24	36.4	4.8	1329	1 US-08-183-213-4	Sequence 4, Appl
25	36	4.8	477	4 US-08-998-416-132	Sequence 132, Appl
26	36	4.8	2540	2 US-08-231-193A-15	Sequence 15, Appl
27	36	4.8	2540	2 US-08-486-273A-15	Sequence 15, Appl

28	36	4.8	2540	3 US-08-480-474-15	Sequence 15, Appl
29	36	4.8	2540	3 US-08-940-086A-15	Sequence 15, Appl
30	36	4.8	2621	3 US-08-436-332B-11	Sequence 1, Appl
31	36	4.8	2727	3 US-08-436-332B-13	Sequence 13, Appl
32	36	4.8	2835	3 US-08-436-332B-12	Sequence 12, Appl
33	36	4.8	2916	3 US-08-436-332B-11	Sequence 11, Appl
34	36	4.8	3007	2 US-08-231-193A-13	Sequence 13, Appl
35	36	4.8	3007	2 US-08-480-474-33	Sequence 33, Appl
36	36	4.8	3007	3 US-08-480-474-33	Sequence 33, Appl
37	36	4.8	3007	3 US-08-940-086A-33	Sequence 33, Appl
38	36	4.8	3070	2 US-08-231-193A-39	Sequence 39, Appl
39	36	4.8	3070	2 US-08-486-273A-39	Sequence 39, Appl
40	36	4.8	3070	3 US-08-480-474-39	Sequence 39, Appl
41	36	4.8	3070	3 US-08-940-086A-39	Sequence 39, Appl
42	36	4.8	3155	2 US-08-231-193A-13	Sequence 13, Appl
43	36	4.8	3155	2 US-08-486-273A-13	Sequence 13, Appl
44	36	4.8	3155	3 US-08-480-474-13	Sequence 13, Appl
45	36	4.8	3155	3 US-08-940-086A-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-08-056-200-93
Sequence 93, Application US/08056200
Patent No. 5616500

GENERAL INFORMATION:

APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Ji
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fredrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:

LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron

APPLICATION NUMBER: US/08/127 721A
FILING DATE: 27-SEPTEMBER-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/952,802
FILING DATE: 25-SEPTEMBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 6066718ak, Henry P.
REGISTRATION NUMBER: 33,200
REFERENCE/DOCKET NUMBER: 4-19276/A/P2/C1H
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 277-5110
TELEFAX: (908) 277-4306
INFORMATION FOR SEQ ID NO: 17:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/127,721
FILING DATE: 27-SEPTEMBER-1993
APPLICATION NUMBER: US 07/952,802
FILING DATE: 25-SEPTEMBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: NO. 60720354k, Henry P.
REGISTRATION NUMBER: 35,200
REFERENCE/DOCKET NUMBER: 4-19276/A/P2/C1P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 277-5110
TELEFAX: (908) 277-4306

FILING DATE: _____
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA: _____
 APPLICATION NUMBER: 08/127,721
 FILING DATE: 27-SEPTEMBER-1993
 APPLICATION NUMBER: US 07/952,802
 FILING DATE: 25-SEPTEMBER-1992
 ATTORNEY/AGENT INFORMATION: _____
 NAME: NO. 5958708ak, Henry P.
 REGISTRATION NUMBER: 33,200
 REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIFH
 TELECOMMUNICATION INFORMATION: _____
 TELEPHONE: (908) 277-5110

FILING DATE: 424
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/127,721
 FILING DATE: 27-SEPTEMBER-1993
 APPLICATION NUMBER: US 07/952,802
 FILING DATE: 25-SEPTEMBER-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: NO. 6072035ak, Henry P.
 REGISTRATION NUMBER: 33,200
 REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIF
 TELECOMMUNICATION INFORMATION:
 TELEPHONE:

TELEFAX: (908) 277-4306
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 22..447
NAME/KEY: mat_peptide
LOCATION: 79..447
OTHER INFORMATION: /product= "heavy chain variable"
OTHER INFORMATION: region C21-H3"
US-08-485-246A-13

Query Match 5.2%; Score 39.2; DB 3; Length 468;
Best Local Similarity 52.4%; Pred. No. 0.016;
Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 32 ggaacgaatcgcgcggaggggacccaccccttcacgtctcctatcgagggcaacatcgaca 91
DB 278 GGGCCACCTTACCGCGGACACCAACACACCGCCTACATGAGAGTGAAGAGCTGA 337
QY 92 ggggaagacacgatttgaaacacttcgagagaaacagaaacgacatttgccctgca 151
DB 338 GAGCGAGGAGACCGCGGTGTACTACTGCGCCGCTTACAGCCACTTACAGGCGACCACT 397
QY 152 cgaagccgtcgagaaatgagcgaactcaactcaacgggtaactctg 195
DB 398 ACGACTACTTGCAGCTACTGGGCGACGGCACCTGTGTGACCGTG 441

RESULT 10

US-08-476-176B-15
Sequence 15, Application US/08476176B
Patent No. 5958708
GENERAL INFORMATION:
APPLICANT: Hardman, No. 5958708man
APPLICANT: Kolbinger, Frank
APPLICANT: Saldanha, Jose
TITLE OF INVENTION: Reshaped monoclonal antibodies against an
NUMBER OF SEQUENCES: 55
TITLE OF INVENTION: Immunoglobulin isotype
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5958708artis Patent Department
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936-1080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,176B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/127,721
FILING DATE: 27-SEPTEMBER-1993
APPLICATION NUMBER: US 07/952,802
FILING DATE: 25-SEPTEMBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5958708ak, Henry P.
REGISTRATION NUMBER: 33,200
REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIP
TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 277-5110
TELEFAX: (908) 277-4306
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 22..447
NAME/KEY: mat_peptide
LOCATION: 79..447
OTHER INFORMATION: /product= "heavy chain variable"
OTHER INFORMATION: region C21-Hay1"
US-08-476-176B-15

Query Match 5.1%; Score 38.2; DB 2; Length 467;
Best Local Similarity 52.1%; Pred. No. 0.032;
Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 33 gaccgaatcgcgcggaggggacccaccccttcacgtctcctatcgagggcaacatcgagcag 92
DB 279 GGGCCACCTTACCGCGGACACCAACACACCGCCTACATGAGAGTGAAGAGCTGA 338
QY 93 cgggaagacacgatttgaaacacttcgagagaaacagaaacgacatttgccctgca 152
DB 339 GAGCGAGGAGACCGCGGTGTACTACTGCGCCGCTTACAGCCACTTACAGGCGACCACTA 398
QY 153 cgaagccgtcgagaaatgagcgaactcaactcaacgggtaactctg 195
DB 399 CGACTACTTGCAGCTACTGGGCGACGGCACCTGTGTGACCGTG 441

RESULT 11

US-08-127-721A-15
Sequence 15, Application US/08127721A
Patent No. 6066718
GENERAL INFORMATION:
APPLICANT: Hardman, No. 6066718man
APPLICANT: Kolbinger, Frank
APPLICANT: Saldanha, Jose
TITLE OF INVENTION: Reshaped monoclonal antibodies against an
NUMBER OF SEQUENCES: 55
TITLE OF INVENTION: Immunoglobulin isotype
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6066718artis Patent and Trademark Department
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936-1080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,721A
FILING DATE: 27-SEPTEMBER-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/952,802
FILING DATE: 25-SEPTEMBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 6066718ak, Henry P.
REGISTRATION NUMBER: 33,200
REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 277-5110

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIP

TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 277-5110
TELEFAX: (908) 277-4306
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 22..447
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 79..447
OTHER INFORMATION: /product="heavy chain variable
OTHER INFORMATION: region C21-H1"
US-08-476-176B-11

Query Match 5.1%; Score 38.2; DB 2; Length 468;
Best Local Similarity 52.1%; Pred. No. 0.032;
Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 33 gaccaagtaacgagagagcaccaccccttcacgcgtccatcgagggcaacatcgag 92
DB 279 GGCACCTTCCACCGCCGACACCAACACCCCTTCAATGAGCTGAGCAGCTGAC 338
QY 93 cgggaagaccagctatttgaaccacttcgagaagtaacaagacatctgctgac 152
DB 339 CAGGAGGAGACACCGCCGTGTACTACTGCGCCAGGTTCAACCCACTTCAGSGGACGACACTA 398
QY 153 cgagcccgctcgaagaatgagcgaacgtaacgggtaactcg 195
DB 399 CGACTACTTGACTACTGCGGCGGACACCCCTGTGACCGTG 441

RESULT 14
US-08-127-721A-11
Sequence 11, Application US/08127721A
Patent No. 6066718
GENERAL INFORMATION:
APPLICANT: Hardman, No. 6066718man
APPLICANT: Kolbinger, Frank
APPLICANT: Saldanha, Jose
TITLE OF INVENTION: Reshaped monoclonal antibodies against an
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6066718artis Patent and Trademark Department
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936-1080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,721A
FILING DATE: 27-SEPTEMBER-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/952,802
FILING DATE: 25-SEPTEMBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 6066718ak, Henry P.
REGISTRATION NUMBER: 33,200
REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIP
TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 277-5110
TELEFAX: (908) 277-4306
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 22..447
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 79..447
OTHER INFORMATION: /product="heavy chain variable
OTHER INFORMATION: region C21-H1"
US-08-127-721A-11

Query Match 5.1%; Score 38.2; DB 3; Length 468;
Best Local Similarity 52.1%; Pred. No. 0.032;
Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 33 gaccaagtaacgagagagcaccaccccttcacgcgtccatcgagggcaacatcgag 92
DB 279 GGCACCTTCCACCGCCGACACCAACACCCCTTCAATGAGCTGAGCAGCTGAC 338
QY 93 cgggaagaccagctatttgaaccacttcgagaagtaacaagacatctgctgac 152
DB 339 CAGGAGGAGACACCGCCGTGTACTACTGCGCCAGGTTCAACCCACTTCAGSGGACGACACTA 398
QY 153 cgagcccgctcgaagaatgagcgaacgtaacgggtaactcg 195
DB 399 CGACTACTTGACTACTGCGGCGGACACCCCTGTGACCGTG 441

RESULT 15
US-08-485-246A-11
Sequence 11, Application US/08485246A
Patent No. 6072035
GENERAL INFORMATION:
APPLICANT: Hardman, No. 6072035man
APPLICANT: Kolbinger, Frank
APPLICANT: Saldanha, Jose
TITLE OF INVENTION: Reshaped monoclonal antibodies against an
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6072035artis Patent Department
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936-1080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,246A
FILING DATE: 27-SEPTEMBER-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/127,721
FILING DATE: 25-SEPTEMBER-1992
APPLICATION NUMBER: US 07/952,802
FILING DATE: 25-SEPTEMBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 6072035ak, Henry P.
REGISTRATION NUMBER: 33,200
REFERENCE/DOCKET NUMBER: 4-19276/A/P2/CIP


```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 277-5110
; TELEFAX: (908) 277-4306
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 468 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 22..447
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 79..447
; OTHER INFORMATION: /product="heavy chain variable
; OTHER INFORMATION: region C21-H1"
US-08-485-246A-11

```

```

Query Match      5.1%; Score 38.2; DB 3; Length 468;
Best Local Similarity 52.1%; Pred. No. 0.032;
Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

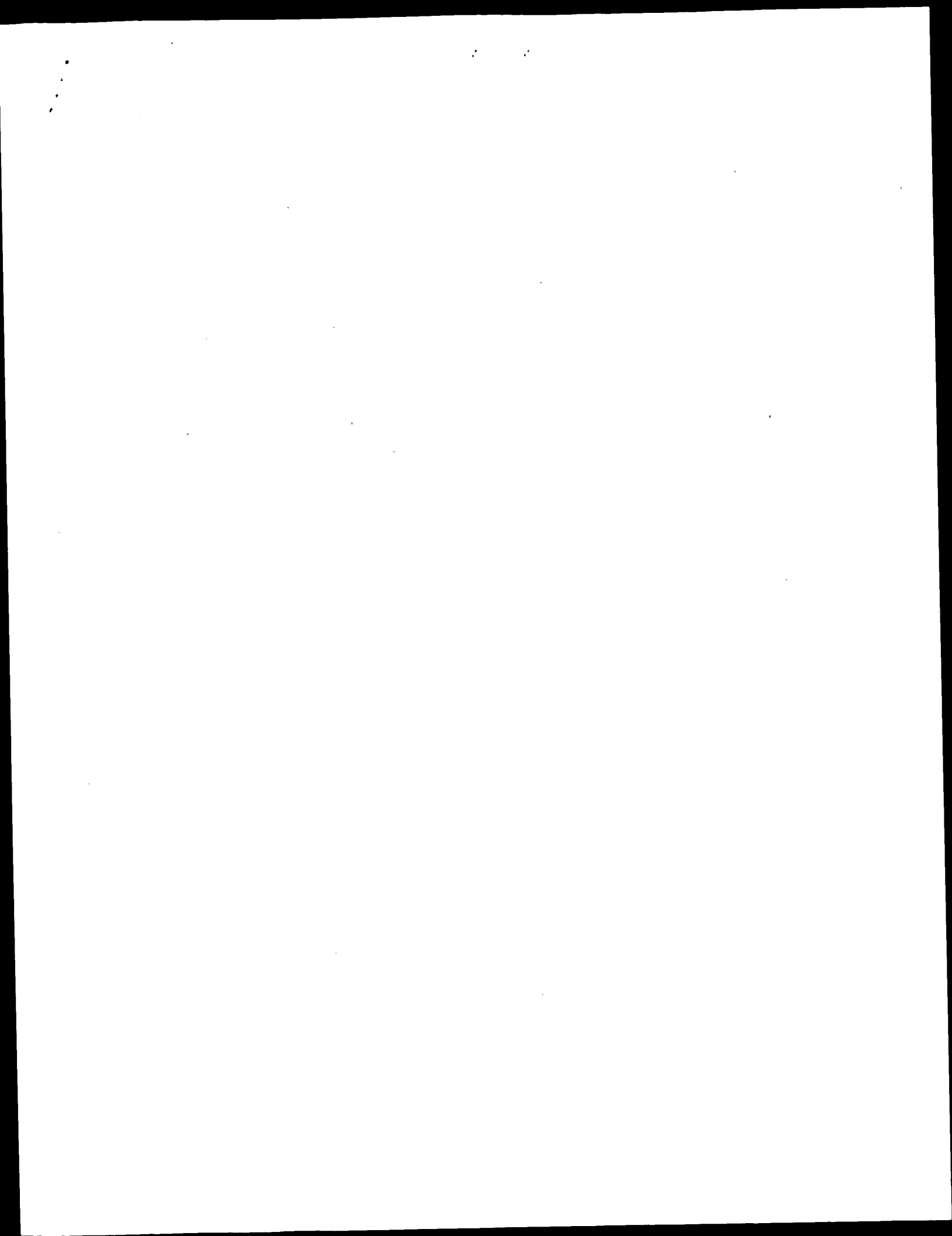
QY 33 gaccagaagcgcgagggcaccacccctccacgcgtccatcgagggcaacatcgagcag 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 GCCACCTTACGCCGCCGACACACACACACACCGCCTACATGAGCTGAGCAGCCTGAC 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 93 cgggaagaccacgatttgaccacattcgagaagtaagaagacatttgctctgac 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 339 CAGCGAGGACACCGCCGCTGTACTACTGCGCCAGGTTCAAGCCTTCAAGCGGACCAACTA 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 cgaagccgtcgagaagtgcgcaagtcacaagggtaaatctg 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 CGACTACTTGACTACTGCGGCCAGGCGACCCCTGTGACCGTG 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: October 28, 2001, 06:24:06
 Job time: 2467 sec



XX The present sequence represents a Drosophila melanogaster
 CC deoxynucleoside kinase DNA sequence. The kinase remains stable
 CC during nucleoside monophosphate synthesis without the presence of
 CC stabilizing reagents. Deoxynucleoside kinases catalyse the
 CC phosphorylation of deoxynucleosides or deoxynucleosides. The enzyme
 CC therefore has an important role in the salvage pathway of nucleotide
 CC metabolism. The recombinant deoxynucleoside kinase is useful for
 CC the phosphorylation of a nucleoside to produce a nucleoside
 CC monophosphate.

XX Sequence 753 BP; 190 A; 215 C; 212 G; 135 T; 1 other;

Query Match 99.9%; Score 752.6; DB 21; Length 753;
 Best Local Similarity 100.0%; Pred. No. 4.4e-197;
 Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcgaagcagcagcctctgtccgaagggggaacgaagtcagcagcagccagccc 60
 DB 1 atggcgaagcagcagcctctgtccgaagggggaacgaagtcagcagcagccagccc 60
 QY 61 ttacccctccatccgagggggaacatcgagcagcagcagcagcagcagcagcagc 120
 DB 61 ttacccctccatccgagggggaacatcgagcagcagcagcagcagcagcagcagc 120
 QY 121 gagagtaacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 180
 DB 121 gagagtaacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 180
 QY 181 aacgggtaaatctgtcggagcagcagcagcagcagcagcagcagcagcagcagc 240
 DB 181 aacgggtaaatctgtcggagcagcagcagcagcagcagcagcagcagcagcagc 240
 QY 241 cagagttatgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
 DB 241 cagagttatgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
 QY 301 aaataaaggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
 DB 301 aaataaaggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
 QY 361 aacggcgtcggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
 DB 361 aacggcgtcggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
 QY 421 gagtcattcaactgtcagcagcagcagcagcagcagcagcagcagcagcagc 480
 DB 421 gagtcattcaactgtcagcagcagcagcagcagcagcagcagcagcagcagc 480
 QY 481 tagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
 DB 481 tagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
 QY 541 cagagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
 DB 541 cagagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
 QY 601 aagtcctcagtcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
 DB 601 aagtcctcagtcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
 QY 661 tcggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
 DB 661 tcggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
 QY 721 tcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 753
 DB 721 tcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 753

RESULT 2
 AAAl1980

ID AAAl1980 standard; DNA; 792 BP.

XX AAAl1980;

XX 07-AUG-2000 (first entry)

DE D. melanogaster deoxynucleoside kinase DM-dnk DNA.

XX Deoxynucleoside kinase; DM-dnk; fruit fly; nucleoside phosphorylation;
 XX nucleoside monophosphate synthesis; insect; PCR; ds.

XX Drosophila melanogaster.

XX Key Location/Qualifiers

XX CDS 22..774

XX FT /tag= a

XX FT /product= "Dm-dnk"

XX DE19846838-A1.

XX 13-APR-2000.

XX 12-OCT-1998; 98DE-1046838.

XX 12-OCT-1998; 98DE-1046838.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX Munch-Petersen B, Pliskur J, Sondergaard L, Ihlenfeldt H, Klein C;

XX Hagedorn U;

XX WPI; 2000-294113/26.

XX Recombinant Drosophila deoxynucleotide kinase useful for preparing

XX nucleoside monophosphates by phosphorylating nucleosides

XX Disclosure; Fig 5; 16pp; German.

XX This invention describes a novel recombinant kinase (I) obtainable from

XX insect cells, which is stable during nucleoside monophosphate synthesis

XX without the addition of stabilizing sulphydryl reagent or stabilizing

XX proteins and accepts all four natural deoxynucleotides. (I) is useful

XX for phosphorylating nucleosides to prepare nucleoside monophosphates.

XX Nucleoside monophosphates are useful as intermediates for nucleoside

XX triphosphates as polymerase chain reaction (PCR) reagents. This sequence

XX represents a DNA fragment which encodes the Drosophila melanogaster

XX deoxynucleoside kinase, designated DM-dnk, which is described in the

XX method of the invention.

XX Sequence 792 BP; 197 A; 224 C; 224 G; 146 T; 1 other;

XX Query Match 99.9%; Score 752.6; DB 21; Length 792;

XX Best Local Similarity 100.0%; Pred. No. 4.5e-197;

XX Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcgaagcagcagcctctgtccgaagggggaacgaagtcagcagcagccagccc 60
 DB 22 atggcgaagcagcagcctctgtccgaagggggaacgaagtcagcagcagccagccc 81
 QY 61 ttacccctccatccgagggggaacatcgagcagcagcagcagcagcagcagcagc 120
 DB 82 ttacccctccatccgagggggaacatcgagcagcagcagcagcagcagcagcagc 141
 QY 121 gagagtaacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 180
 DB 142 gagagtaacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 201
 QY 181 aacgggtaaatctgtcggagcagcagcagcagcagcagcagcagcagcagcagc 240
 DB 202 aacgggtaaatctgtcggagcagcagcagcagcagcagcagcagcagcagcagc 261
 QY 241 cagagttatgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300

```

Db      262 cagagttatgtaacgtacatgctgacgtgcacacccgcccacacacagaagcta 321
Qy      301 aaaaataargagagcgtccatcttttagcgtcgtctatgtctctgtagaagaatgcagca 360
Db      322 aaaaataargagagcgtccatcttttagcgtcgtctatgtctctgtagaagaatgcagca 381
Qy      361 aacggtcgtctgagacagagcgtacatacgtctgagaggtgtgtacaagtctcgaa 420
Db      382 aacggtcgtctgagacagagcgtacatacgtctgagaggtgtgtacaagtctcgaa 441
Qy      421 gagtcattacactgcagagcgagacctcatatattctgcacacctgcgcgaggtgagc 480
Db      442 gagtcattacactgcagagcgagacctcatatattctgcacacctgcgcgaggtgagc 501
Qy      481 tacgaacgcatccgagcagcgagcgtctgtctgtagagagatcgtgtagcgttaagtaactt 540
Db      502 tacgaacgcatccgagcagcgagcgtctgtctgtagagagatcgtgtagcgttaagtaactt 561
Qy      541 caggaagctgcatgagttgacacagagactggtgtatatacaccagagacgacgacgtctgac 600
Db      562 caggaagctgcatgagttgacacagagactggtgtatatacaccagagacgacgacgtctgac 621
Qy      601 aaggtccctagctctcgtatgacgtcgtatgaaacctggaacaattggcaccgagttaccagcgc 660
Db      622 aaggtccctagctctcgtatgacgtcgtatgaaacctggaacaattggcaccgagttaccagcgc 681
Qy      661 tcgagagacagacatactgcagcaccatctcaagttacacacagcagcgcctgcggttcgtgtg 720
Db      682 tcgagagacagacatactgcagcaccatctcaagttacacacagcagcgcctgcggttcgtgtg 741
Qy      721 tcgcccagaagcgccagagaggtgcgcagataa 753
Db      742 tcgcccagaagcgccagagaggtgcgcagataa 774

RESULT 3
AAAS1375 standard; cDNA; 1113 BP.
ID      AAAS1375 standard; cDNA; 1113 BP.
AC      AAAS1375;
DT      26-SEP-2000 (first entry)
DE      D. melanogaster dNK, multisubstrate deoxyribonucleosidase cDNA.
KW      dNK; multisubstrate deoxyribonucleosidase; kinase; cytosolic;
OS      Drosophila melanogaster.
FH      Key Location/Qualifiers
FT      CDS 113..865
FT      /*tag= a
FT      /product= Multisubstrate_deoxyribonucleosidase
PN      MO200036099-A1.
PD      22-JUN-2000.
PE      10-DEC-1999; 99MO-SE02314.
PR      11-DEC-1998; 98SE-0004298.
PA      (KARL/) KARLSSON A.
PI      Karlsson A, Johansson M;
PX      DR
PX      MPI: 2000-431584/37.
PX      P-PSDB; AAY96812.
PX      New nucleic acid encoding a multisubstrate deoxyribonucleosidase with
PX      homology to a Drosophila melanogaster sequence for activating

```

```

PT      anti-viral and anti-cancer drugs to treat viral infections and cancer
XX
PS      Claim 1; Page 33; 38pp; English.
PX
XX      This cDNA encodes Drosophila melanogaster multisubstrate
CC      deoxyribonucleosidase (dNK). The multisubstrate deoxyribonucleoside
CC      changes nucleoside analogue products into active anti-cancer and
CC      anti-viral drugs by phosphorylation, allowing cancer and viral
CC      infections to be treated. Drosophila melanogaster deoxyribonucleoside
CC      kinase is (unlike human deoxyribonucleoside kinase) a multisubstrate
CC      enzyme and it catalyzes phosphorylation of pyrimidine and purine
CC      nucleoside analogues. The catalytic rates of deoxyribonucleoside and
CC      maximal catalytic rates reported for mammalian enzymes, making it more
CC      useful for activating anti-viral and anti-cancer drugs.
XX
SQ      Sequence 1113 BP; 324 A; 287 C; 272 G; 230 T; 0 other;

Query Match      98.3%; Score 740.2; DB 21; Length 1113;
Best Local Similarity 98.9%; Pred. No. 1,3e-193;
Matches 745; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy      1 atgagcagagcagacatctgtcccgaaaggagacaaagtaacgacgagcagacccccc 60
Db      113 atgagcagagcagacatctgtcccgaaaggagacaaagtaacgacgagcagacccccc 172
Qy      61 tcaaccgtctcatcgaaggcaacacatcgcagcagcgaagaaacacgatttgacacacctc 120
Db      173 ttaaccgtctcatcgaaggcaacacatcgcagcagcgaagaaacacgatttgacacacctc 232
Qy      121 gaaagttacagaagaacacatttgcctcgtcgtacccgagcccgtagaagtggcgcaagtc 180
Db      233 gaaagttacagaagaacacatttgcctcgtcgtacccgagcccgtagaagtggcgcaagtc 292
Qy      181 aacggggttaaatctgtctgagctgagctgatacagaagaatcccaagaatgggacatgccttt 240
Db      293 aacggggttaaatctgtctgagctgagctgatacagaagaatcccaagaatgggacatgccttt 352
Qy      241 cagagttatgtaacgctgacacatgctgcagtcgcacacccgcccaacaaagaagcta 300
Db      353 cagagttatgtaacgctgacacatgctgcagtcgcacacccgcccaacaaagaagcta 412
Qy      301 aaaaataargagagcgtccatttttagcgtcgtctatgtctctgtgtgagaactgcagca 360
Db      413 aaaaataargagagcgtccatttttagcgtcgtctatgtctctgtgtgagaactgcagca 472
Qy      361 aacggtcgtctgagacagagcgtacatacgtctggaagatgtgtacaagtctcgaa 420
Db      473 aacggtcgtctgagacagagcgtacatacgtctggaagatgtgtacaagtctcgaa 532
Qy      421 gagtcattacactgcagagcgagacctcatatattctgcacacctgcgcgaggtgagc 480
Db      533 gagtcattacactgcagagcgagacctcatatattctgcacacctgcgcgaggtgagc 592
Qy      481 tacgaacgcatccgagcagcgagcgtctgtctgtagagagatcgtgtgcgttaagtaactt 540
Db      593 tacgaacgcatccgagcagcgagcgtctgtctgtagagagatcgtgtgcgttaagtaactt 652
Qy      541 caggaagctgcatgagttgacacagagactggtgtatatacaccagagacgacgacgtctgac 600
Db      653 caggaagctgcatgagttgacacagagactggtgtatatacaccagagacgacgacgtctgac 712
Qy      601 aaggtccctagctctcgtatgacgtcgtatgaaaccttggaaaacattggcaccgagttaccagcgc 660
Db      713 aaggtccctagctctcgtatgacgtcgtatgaaaccttggaaaacattggcaccgagttaccagcgc 772
Qy      661 tcgagagacagacatactgcagcaccatctcaagtaacacacagcgcctgcggttcgtgtg 720
Db      773 tcgagagacagacatactgcagcaccatctcaagtaacacacagcgcctgcggttcgtgtg 832
Qy      721 tcgcccagaagcgccagagaggtgcgcagataa 753

```


PD 09-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-N200040.
 XX
 PR 29-MAR-1996; 96NZ-0286284.
 XX
 PA (UYOT-) UNIV OTAGO.
 XX
 PI Lytle DJ, Robinson AJ;
 XX
 DR WPI: 1997-503116/46.
 XX
 PF Parapoxvirus vectors containing exogenous DNA - comprise parapox
 FT virus, especially orf virus, and exogenous DNA encoding a
 PT polypeptide, e.g. an antigen useful in vaccines
 XX
 PS
 XX
 PS Claim 12: Fig 7, 73pp: English.
 CC This genomic DNA molecule comprises the BamHI F fragment and part
 CC of the BamHI C fragment of the orf virus strain NZ-2 genome
 CC incorporating putative promoter sequences for the Fil, F2L, F4R
 CC and C1R genes as well as intergenic regions suitable as foreign
 CC gene insertion sites. A claimed parapoxvirus vector comprises a
 CC parapox virus (especially an orf virus) containing exogenous DNA,
 CC and host cells (especially bovine or ovine testis cells)
 CC incorporating the vector are also claimed. The vector and
 CC fragments/variants are useful in vaccine production (claimed),
 CC since they can express antigens, useful in medical and veterinary
 CC applications e.g. parasitic disease control in livestock. Particular
 CC antigens include HIV envelope protein, herpes simplex glycoprotein,
 CC antigens of *Escherichia coli*, *Trichostromyces*, *Haemophilus*,
 CC *Osteoblast* and especially *Taenia ovis* 45W, 16 kd, 18 kd antigens.
 CC The foreign gene is preferably incorporated in a non-essential
 CC region of the virus genome (see AT91357-61) under control of a
 CC poxvirus promoter, especially orf virus promoter E1L, Fil or F3R
 CC (see AT91362-64). The orf virus vectors limit the cross-species
 CC infection common with e.g. vaccinia virus vectors, since the virus
 CC has a narrow host range; orf is also less virulent than vaccinia in
 CC man.
 XX
 XX Sequence 4372 BP; 817 A; 1346 C; 1439 G; 770 T; 0 other;
 SQ
 Query Match 6.0%; Score 45.4; DB 18; Length 4372;
 Best Local Similarity 47.6%; Pred. No. 0.0087;
 Matches 130; Conservative 1; Mismatches 142; Indels 0; Gaps 0;

QY 48 ggggacccagccttaccgtctctatcgagggaacatcgagcggaagacacgta 107
 DB 2352 GGGCACCACCTCTACACGGACTTCAATCGCCGAGAACATCCATCTTCGTGCAACT 2293
 QY 108 ttgaaccacttcgagaagatacaagaacatttcctgctcagcagagccgctgaga 167
 DB 2292 GTACAAAGTGTGAGCGATTCACGAGAGACGACATCCGCTGCTTTCGACCATGTGAGCA 2233
 QY 168 gtgagcgaacgttcaacgggttaactgtctgagctgagtacaaagatcccaagaagt 227
 DB 2232 GAACCCCGAGCGAGCGACGCCACGCTGATTCGTGATATCTCTACATCGGAGAACGTT 2173
 QY 228 ggcacatgccttcagagatgtatcgctgaccatgctgcagtgacacagccccaac 287
 DB 2172 CGAGAAAGCAGTGTGAGACGACGCCACCGAGACCTCTCGTGGCGGACGACGACCGGA 2113
 QY 288 caacagaagaataaataaargagcgtccat 320
 DB 2112 CGACACTCGGACAGCATGTAGCAGATGTTCTT 2080

RESULT 6
 AAV22141
 ID AAV22141 standard; cDNA; 36519 BP.
 XX

AC AAV22141;
 XX
 DT 17-AUG-1998 (first entry)
 XX
 DE Chimpanzee adenovirus C68 genomic sequence.
 XX
 XX genomic sequence; viral genome; virus; gene therapy; treatment;
 KW acquired defects; inherited defects; genetic engineering; vector;
 KM in vitro production; recombinant protein; ds.
 XX
 OS Chimpanzee adenovirus C68.
 XX
 FH Key Location/Qualifiers
 FT CDS 480..1521
 FT /tag= a
 FT /note= "E1A gene"
 FT 1560..3956
 FT /tag= b
 FT /note= "E1B gene"
 FT complement(23370..21787)
 FT /tag= c
 FT /note= "E2A gene"
 FT complement(10346..3957)
 FT /tag= d
 FT /note= "E2B gene"
 FT 26806..31877
 FT /tag= e
 FT /note= "E3 gene"
 FT complement(36193..33486)
 FT /tag= f
 FT /note= "E4 gene"
 FT 10823..13817
 FT /tag= g
 FT /note= "L1 gene"
 FT 13884..17431
 FT /tag= h
 FT /note= "L2 gene"
 FT 17480..21804
 FT /tag= i
 FT /note= "L3 gene"
 FT 23399..27439
 FT /tag= j
 FT /note= "L4 gene"
 FT 32134..33502
 FT /tag= k
 FT /note= "L5 gene"
 XX
 XX W09810087-A1.
 XX
 XX 12-MAR-1998.
 XX
 PF 04-SEP-1997; 97WO-US15694.
 XX
 PR 06-SEP-1996; 96US-0024700.
 XX
 PA (UYPE-) UNIV PENNSYLVANIA.
 XX
 PI Farina SF, Fisher KJ, Wilson JM;
 XX
 DR WPI: 1998-193635/17.
 XX
 PT Chimpanzee adenovirus vector - useful in gene therapy and genetic
 FT engineering in general
 XX
 PS Disclosure: Pages 70-89; 116pp: English.
 XX
 CC The sequence is that of the chimpanzee adenovirus C68. It may be
 CC used in the production of a vector comprising it and a selected
 CC heterologous gene operatively linked to regulatory sequences
 CC directing its expression in a heterologous cell.
 CC Such recombinant adenoviruses are useful as vectors in gene
 CC therapy and genetic engineering in general, especially for
 CC the treatment of acquired or genetically inherited defects.

AAV72046
 ID AAV72046 standard; DNA; 1929 BP.
 AC AAV72046;
 XX
 XX
 DT 10-MAY-1999 (first entry)
 DE Clome PNG4/ASB7VH-IGG2CH1/CPG2 R6 DNA.
 XX
 XX Conjugate; cell targeting; cytotoxic drug; produg therapy system;
 KM produg-converting enzyme; cell surface antigen; treatment; cancer;
 KM inflammation; rheumatoid arthritis; antibody; ss.
 XX
 OS Synthetic.
 XX
 PN WO9851787-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 05-MAY-1998; 98WO-GB01294.
 XX
 PR 10-MAY-1997; 97GB-0009421.
 XX
 PA (ZENE) ZENECA LTD.
 XX
 PI Blakey DC, Emery SC;
 XX
 DR WPI; 1999-059700/05.
 DR P-PSDB; AAM82739.
 XX
 PT New gene construct expressing conjugate of targeting agent and
 PT produg-converting enzyme - useful for, e.g. targeted production of
 PT cytotoxic drug in vivo, especially for treatment of cancer
 XX
 XX Example 1e; Page 67-68; 100pp; English.
 PS
 XX This sequence is used in a method for obtaining a novel gene construct
 CC (A) which expresses, in cells of a mammal, a conjugate (B) of a
 CC cell-targeting group (I) and a heterologous produg-converting enzyme
 CC (II), and (B) is directed to leave the cell for selective localisation at
 CC a cell surface antigen (Ag) recognised by (I). Delivery of (A) to a
 CC target site, then administration of (III) is used for targeted release of
 CC cytotoxic drug, specifically for treating cancer but also inflammation
 CC such as rheumatoid arthritis. In situ generation of the targeting
 CC antibody increases selectivity, reducing side effects at normal tissue.
 CC The method is applicable to any antibody-directed enzyme produg therapy
 CC system.
 CC
 XX Sequence 1929 BP; 409 A; 603 C; 597 G; 320 T; 0 other;
 SQ

Query Match 5.3%; Score 39.8; DB 20; Length 1929;
 Best Local Similarity 47.0%; Pred. No. 0.22;
 Matches 119; Conservative 1; Mismatches 133; Indels 0; Gaps 0;
 OY 150 gaccgaagccgcgcgaagtgcgcaacgtcaacggggtlaaactgtgtagctgtagtga 209
 DB 1383 gcccgagcgcggcggaacgcgtggtgtcaggtcttcgaactcgtgctgcgcaagatgaa 1442
 OY 210 caaagatcccaagaagtggcgcgtcccttcagaagtatttcacgcgtgacacatgctga 269
 DB 1443 catcgacgaagaagcgaagaaactgcgttcagtggaaccatgcgaagcgcgcaaggt 1502
 OY 270 gtccgacacggcccaacgaagaagctaaataatargagacgctccatttttagcgc 329
 DB 1503 ctggaacatcatcccgccagcgccacgctgacgacgcgaactgtgctacgcgcaagga 1562
 OY 330 tcgctattcgtcgtggaagacatgcagacgaacgcgtgcgtgcggaagcgatgtataa 389
 DB 1563 ggaacttcgacgcgcgcacatgaagacgtggaagagcgcgcgacgacgaagaagctgcccga 1622
 OY 390 tacgctggaagag 402

DB 1623 ggcgcagctgaag 1635

RESULT 14

AAV72075
 ID AAV72075 standard; DNA; 1998 BP.

AC AAV72075;
 XX
 XX
 DT 10-MAY-1999 (first entry)
 DE Fusion protein (806.077 scFv-CPG2) 2 DNA.
 XX
 XX
 XX Conjugate; cell targeting; cytotoxic drug; plasmid; fusion protein;
 KM produg-converting enzyme; cell surface antigen; treatment; cancer;
 KM inflammation; rheumatoid arthritis; antibody; produg therapy system; ss.
 XX
 OS Synthetic.
 XX
 PN WO9851787-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 05-MAY-1998; 98WO-GB01294.
 XX
 PR 10-MAY-1997; 97GB-0009421.
 XX
 PA (ZENE) ZENECA LTD.
 XX
 PI Blakey DC, Emery SC;
 XX
 DR WPI; 1999-059700/05.
 DR P-PSDB; AAM82745.
 XX
 PT New gene construct expressing conjugate of targeting agent and
 PT produg-converting enzyme - useful for, e.g. targeted production of
 PT cytotoxic drug in vivo, especially for treatment of cancer
 XX
 XX Example 17; Page 87; 100pp; English.
 PS
 XX This sequence is used in a method for obtaining a novel gene construct
 CC (A) which expresses, in cells of a mammal, a conjugate (B) of a
 CC cell-targeting group (I) and a heterologous produg-converting enzyme
 CC (II), and (B) is directed to leave the cell for selective localisation
 CC at a cell surface antigen (Ag) recognised by (I). Delivery of (A) to a
 CC target site, then administration of (III) is used for targeted release of
 CC cytotoxic drug, specifically for treating cancer but also inflammation
 CC such as rheumatoid arthritis. In situ generation of the targeting
 CC antibody increases selectivity, reducing side effects at normal tissue.
 CC The method is applicable to any antibody-directed enzyme produg therapy
 CC system.
 CC
 XX Sequence 1998 BP; 420 A; 601 C; 628 G; 349 T; 0 other;
 SQ

Query Match 5.3%; Score 39.8; DB 20; Length 1998;
 Best Local Similarity 47.0%; Pred. No. 0.23;
 Matches 119; Conservative 1; Mismatches 133; Indels 0; Gaps 0;
 OY 150 gaccgaagccgcgcgaagtgcgcaacgtcaacggggtlaaactgtgtagctgtagtga 209
 DB 1452 gcccgagcgcggcggaacgcgtggtgtcaggtcttcgaactcgtgctgcgcaagatgaa 1511
 OY 210 caaagatcccaagaagtggcgcgtcccttcagaagtatttcacgcgtgacacatgctga 269
 DB 1512 catcgacgaagaagcgaagaaactgcgttcagtggaaccatgcgaagcgcgcaaggt 1571
 OY 270 gtccgacacggcccaacgaagaagctaaataatargagacgctccatttttagcgc 329
 DB 1572 ctggaacatcatcccgccagcgccacgctgacgacgcgaactgtgctacgcgcaagga 1631
 OY 330 tcgctattcgtcgtggaagacatgcagacgaacgcgtgcgtgcggaagcgatgtataa 389

Db 1632 ggaactcgaagccgcatgaagaagcgtgaaagcgcgcagcagaagaagctgcccga 1691
 QY 390 taagctggaagag 402
 Db 1692 ggcgcagcgtgaag 1704

RESULT 15
 AAV72059
 ID AAV72059 standard; DNA: 2019 BP.
 AC AAV72059;

DT 10-MAY-1999 (first entry)

DE Plasmid pNG4/55.1scfv/CPG2 R6 DNA.

KN Conjugate; cell targeting; cytotoxic drug; prodrug-converting enzyme;
 cell surface antigen; treatment; cancer; inflammation; antibody;
 rheumatoid arthritis; prodrug therapy system; ss.

OS Synthetic.

PN WO9851787-A2.

PD 19-NOV-1998.

PE 05-MAY-1998; 98WO-GB01294.

PR 10-MAY-1997; 97GB-0009421.

PA (ZENE) ZENECA LTD.

PI Blakey DC, Emery SC;

DR WPI: 1999-059700/05.

DR P-PSDB: AAW82742.

PT New gene construct expressing conjugate of targeting agent and
 prodrug-converting enzyme - useful for, e.g. targeted production of
 cytotoxic drug in vivo, especially for treatment of cancer

PS Example 14; Page 77; 100pp; English.

CC This sequence is used in a method for obtaining a novel gene construct
 (A) which expresses, in cells of a mammal, a conjugate (B) of a
 cell-targeting group (I) and a heterologous prodrug-converting enzyme
 (II), and (B) is directed to leave the cell for selective localisation at
 a cell surface antigen (Ag) recognised by (I). Delivery of (A) to a
 target site, then administration of (II) is used for targeted release of
 cytotoxic drug, specifically for treating cancer but also inflammation
 such as rheumatoid arthritis. In situ generation of the targeting
 CC antibody increases selectivity, reducing side effects at normal tissue.
 CC The method is applicable to any antibody-directed enzyme prodrug therapy
 system.

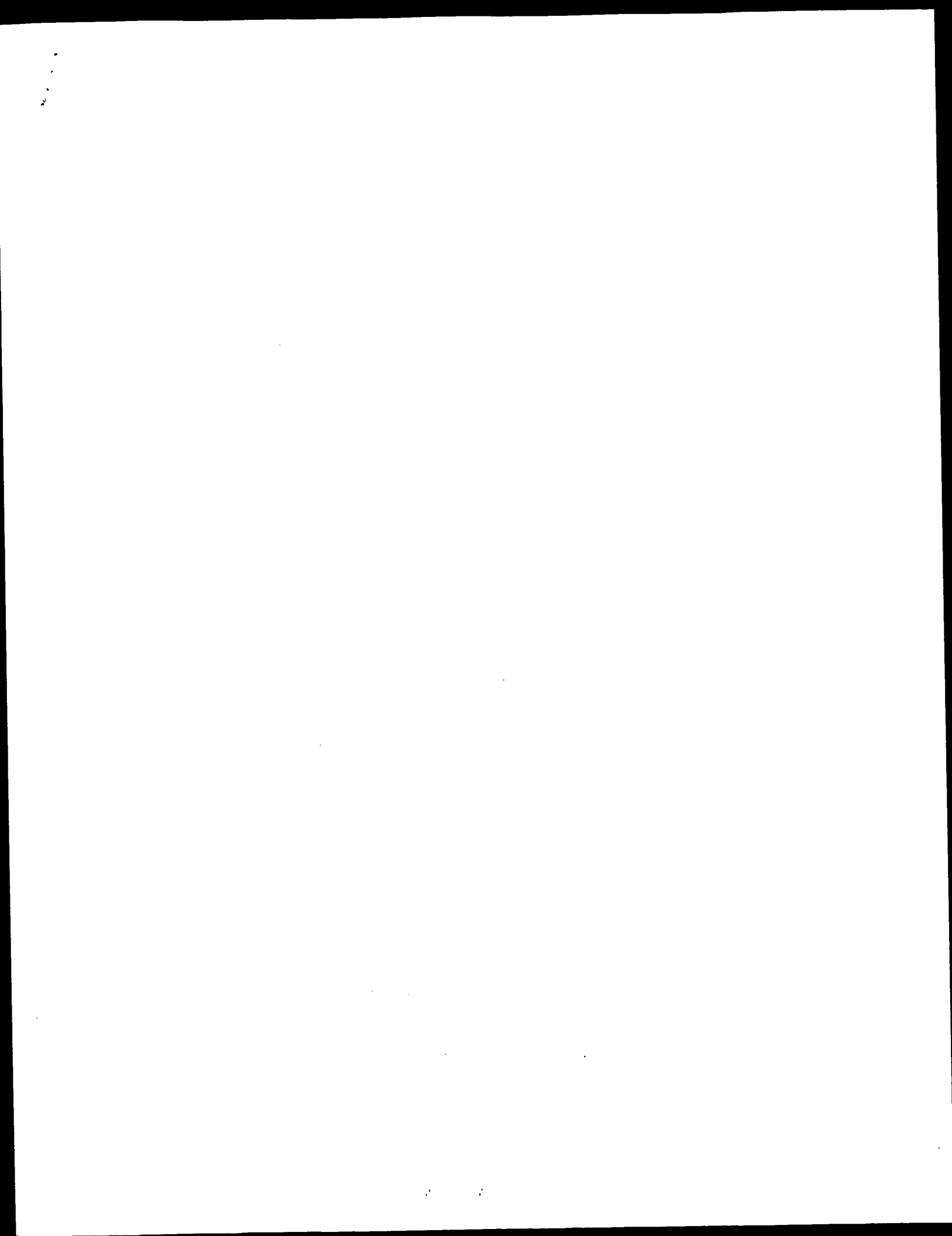
SQ Sequence 2019 BP; 437 A; 604 C; 637 G; 341 T; 0 other;

Query Match 5.3%; Score 39.8; DB 20; Length 2019;
 Best Local Similarity 47.0%; Pred. No. 0.23;
 Matches 119; Conservative 1; Mismatches 133; Indels 0; Gaps 0;

QY 150 gacgcgagccgctcgagaagtgcgcaacggttaattgctgagcgtgatgta 209
 Db 1473 gccgcagctgggctggaacgcgtgctggaagcttcgaacctgctgcgcacagatgaa 1532
 QY 210 caaagatcccaagaagtggccatgaccttcagaagtatgtcacgctgaacatgctgca 269
 Db 1533 catcgacgacgaagcgcaagaaacctgcttcacgtgacacatcgcaagcgccagcagt 1592
 QY 270 gtgcacacgcgcgcccaacacgaagaactaaataargagagctccatttttagcgc 329

Db 1593 ctgnaacatcatcccccagccagcgcagcgtgaacgcgcagctgctacgcgcgcaacga 1652
 QY 330 tcgctatgcttcgctggggaacatgcaacgaacgcgtcgtgagcagggcatgtacaa 389
 Db 1653 ggaactcgaagccgcacatgaagaagcgtggaagagcgcgcagcagaagaagctgcccga 1712
 QY 390 taagctggaagag 402
 Db 1713 ggcgcagcgtgaag 1725

Search completed: October 28, 2001, 06:26:04
 Job time: 2565 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 04:34:04 ; Search time 1077.11 Seconds
(without alignments)
10813.394 Million cell updates/sec

Title: us-09-416-579a-1

Perfect score: 753
Sequence: 1 atggcggagcgacgacatcctg.....gccagagggtcgcacagataa 753

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_om:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_p11:*
- 13: gb_p12:*
- 14: gb_p13:*
- 15: gb_p14:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_higo_hum:*
- 20: em_higo_inv:*
- 21: em_higo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_v1:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_v11:*

59: gb_v12:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_prl:*

86: gb_prl:*

87: gb_prl:*

88: gb_prl:*

89: gb_prl:*

90: gb_prl:*

91: gb_prl:*

92: gb_prl:*

93: gb_prl:*

94: gb_prl:*

95: gb_prl:*

96: gb_prl:*

97: gb_prl:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	745	98.9	1121	5	AF045610
2	742	98.5	750	6	DME18048
3	592	78.6	1246	5	AF185268
4	592	78.6	50829	65	AC017759
5	592	78.6	232380	5	AE003724
6	538.8	71.6	157133	60	AC007814
7	290.4	38.6	102657	60	AC007891
8	186.8	24.8	747	5	AF226281

9	127.2	16.9	1530	94	AF105217	Mus muscu
10	125.6	16.7	2385	94	AJ249341	Mus muscu
11	116.4	15.5	1953	94	HSMTK	Homo sapien
12	113.4	15.1	3060	85	AB046005	Macaca fa
13	113.2	15.0	1860	97	HSU77088	Human thym
14	113.2	15.0	1860	97	HSU80628	Human thym
15	49.2	6.5	1040	94	MMU133749	Mus muscu
16	49.2	6.5	1050	94	MMU90524	Mus muscu
17	49.2	6.5	1071	94	MMU133750	Mus muscu
18	48.4	6.2	2812	94	MMDEOKI	X77731 M.musculus
19	46.8	6.2	9551	94	AR076233	Sequence 93
20	46.8	6.2	9551	10	139845	Sequence 93
21	46.8	6.2	9551	97	HUMTRHYAL	Human trich
22	46.4	6.2	46166	3	SC6F11	AL53864 Streptomy
23	45.4	6.0	1496	14	CRENOLM	S66412 C.reinhardt
24	45.4	6.0	3348	59	S62819	S62819 F2L-putativ
25	45.4	6.0	29924	59	SC5E9	AL446003 Streptomy
26	45.4	6.0	134226	59	IHICG	M75136 Ictulurid h
27	45.2	6.0	950	95	RATDCK	L33899 Rattus norv
28	45.2	6.0	1100	97	HSU41668	U41668 Human deoxy
29	43.2	5.7	288539	58	AF198100	AF198100 Fowlpox v
30	42.2	5.6	2332	5	AF184664	AF184664 Drosophil
31	42.2	5.6	167596	63	AC013177	AC013177 Drosophil
32	42.2	5.6	183007	4	AC007257	AC007257 Drosophil
33	42.2	5.6	260624	4	AE003625	AE003625 Drosophil
34	41.8	5.6	1025	93	HSDDK	X97386 H.sapiens m
35	41.4	5.5	31360	3	SC9C7	AL035161 Streptomy
36	41.4	5.5	11106	1	AE005891	AE005891 Caulobact
37	41.2	5.5	13800	9	SC7B7	AL009199 Streptomy
38	41.2	5.5	36519	9	AR101859	AR101859 Sequence
39	41.2	5.5	41173	3	SC1A1	AL132648 Streptomy
40	40	5.3	30561	9	AX044418	AL013350 Streptomy
41	39.8	5.3	1180	9	AX044418	AX044418 Sequence
42	39.8	5.3	1236	9	AB3189	AB3189 Sequence 9
43	39.8	5.3	1257	9	AX044413	AX044413 Sequence
44	39.8	5.3	1926	9	AB3205	AB3205 Sequence 25
45	39.8	5.3	1929	9	AB3195	AB3195 Sequence 15

ALIGNMENTS

RESULT 1						
LOCUS AF045610	1121 bp	mRNA	INV	24-FEB-2000		
DEFINITION Drosophila melanogaster deoxyribonuclease kinase (Dm-dnk) mRNA,						
complete cds.						
ACCESSION AF045610						
VERSION AF045610.2	GI:7025918					
KEYWORDS						
SOURCE						
ORGANISM						
	fruit fly,					
	Drosophila melanogaster,					
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;					
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE						
AUTHORS	Johansson, M., van Rompay, A.R., Degreve, B., Balzarini, J. and					
	Karlsson, A.					
TITLE	Cloning and characterization of the multisubstrate					
JOURNAL	deoxyribonuclease kinase of the Drosophila melanogaster					
MEDLINE	J. Biol. Chem. 274 (34), 23814-23819 (1999)					
REFERENCE						
AUTHORS	Johansson, M. and Karlsson, A.					
TITLE	2 (bases 1 to 1121)					
JOURNAL	Direct Submission					
	Submitted (02-FEB-1998) Clinical Virology, Karolinska Institute,					
	Huddinge Hospital, Stockholm S-141 86, Sweden					
	3 (bases 1 to 1121)					
REFERENCE						
AUTHORS	Johansson, M. and Karlsson, A.					
TITLE	Direct Submission					
JOURNAL	Submitted (24-FEB-2000) Clinical Virology, Karolinska Institute,					
	Huddinge Hospital, Stockholm S-141 86, Sweden					
REMARK	Sequence update by submitter					

COMMENT On Feb 24, 2000 this sequence version replaced gi:5708379.

FEATURES

source location/Qualifiers

1..1121

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/dev_stage="embryo"

/note="RNA from Drosophila melanogaster embryo BDGP Drosophila EST project clone LD15983"

1..1121

/gene="Dm-dnk"

/gene="125..877"

/note="Dm-dnk"

/note="thymidine kinase"

/product="deoxyribonuclease kinase"

/protein_id="AAD47355.2"

/db_xref="GI:7025919"

/translation="MAEASCARQKTYAEGTOPFYLIIGNIGSKTYYLHNPEKXNDICLLEPEVEKRVNNGVNLLELMYKDPKRWMPFQSYVITLMLQSHAPTKLKI MRSITFSARYCEVENRRNGSLDEQMYWLEEMYKTIESTIHVQADLIYLRSPVEA YRIIRRARSESCVPLKYLDELHELDHEDMLIHORRPOCKVILVDLNLNENIGTEY QNSESTFPAISNOOPSPVLVSPSKRORVAR"

BASE COUNT 323 a 290 c 274 g 234 t

ORIGIN

Query Match 98.9%; Score 745; DB 5; Length 1121;

Best Local Similarity 99.3%; Pred. No. 1.3e-158;

Matches 748; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 atggcggagcagcagcattctgtccggaagggaacgaagtaacgcgagggaccaccc 60

125 ATGGCGGAGCAGCAGCATTCTGTCCGGAAGGAGCAGCAGCAGCAGCAGCAGCAGC 184

61 ttacacctccctcaccgagggcaacatggcaggggaagcaacgatttgacaacttc 120

185 TTACACCTCCCTCACCAGGGCAACATGGCAGGGGAGGAGGAGGAGGAGGAGGAGG 244

121 gagaagtaacaagcagcatttgcctgtcgtaccgagcccgctcgagaagtggcgaagtc 180

245 GAGAAGTACAGAGCAGCAGCATTGCTGCTGACGAGCCGCTGAGAGAGGAGGAGGAGG 304

181 aacggggttaaatctgtcgtgagcgtatgtacaagaatcccaagaagtggcccttc 240

305 AACGGGTAAATCTGCTGAGCCTGATGTACAAAGATCCCAAGAGTGGCCATGCCCTTT 364

241 cagagttatgtcaagcgtacacatgtcagtcgacacacgcccacaaacgaagaagta 300

365 CAGAGTTATGTCAAGCCTGACCATGCTGACGAGCAGCAGCAGCAGCAGCAGCAGCTTA 424

301 aaaaataaraggagcgtccatttttaagcgtcgtcgtatgtctgtgtgagaacatggagcga 360

425 AAAATAATAGGAGCGCTCCATTTTATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484

361 aacgctcgtcgtgagcagcagcattacacacgctgtgaggaagtgtacaagtcatcgaa 420

485 AACGCTGCTGCTGAGCAGCAGCAGTGTACAAATACGCTGAGGAGAGTGTCAAGTTCAATCGAA 544

421 gattccattcaacctgcaagcagcagcattacatataatcgtgcacacctgcgcgaggtggcg 480

545 GATTCATTCAACGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 604

481 tacgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540

605 TACGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 664

541 caggagcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600

665 CAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 724

601 aaggtcctagtcctcagatgcgaatcgtgaacctgtgaacaaatctggcagcagcagcagcagc 660

725 AAGGTCCTAGTCCTCAGATGCAGATCTGAACCTGTGAACCATTTGGACCGAGTACACGAGC 784

QY 661 tcggaagacagcatatcgcagccatctcaagtaacacagacccctgcggttcgtgtg 720
 Db 785 TCGGAGACAGCATATTCGACCCCATCTCAAGTAACCAACAGCCCTCGCGTTCGTG 844
 QY 721 tcgccaagacagcccgagaggttcgcagataa 753
 Db 845 TCGCCAGCAAGCCGACAGAGGTCGCGCAGATAA 877

RESULT 2
 LOCUS DME18048 750 bp mRNA INV 03-MAR-2000
 DEFINITION Drosophila melanogaster mRNA for deoxyribonucleoside kinase, partial.
 ACCESSION Y18048
 VERSION Y18048.1 GI:4741227
 KEYWORDS deoxynucleoside kinase; dnk gene.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 750)
 Munch-Petersen, B., Knecht, W., Lenz, C., Sondergaard, L. and Piskur, J.
 Functional expression of a multisubstrate deoxyribonucleoside
 kinase from Drosophila melanogaster and its C-terminal deletion
 mutants
 J. Biol. Chem. 275 (9), 6673-6679 (2000)
 MEDLINE 20158988
 REFERENCE 2 (bases 1 to 750)
 Piskur, J.
 AUTHORS Direct Submission
 TITLE Submitted (01-SEP-1998) J. Piskur, Department of Microbiology,
 JOURNAL Technical University of Denmark, DTU - 301, DK - 2800, Lyngby,
 DENMARK

FEATURES
 source
 1..750
 /location="Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /dev_stage="embryo"
 1..750
 /gene="dnk"
 1..2750
 /gene="dnk"
 /codon_start=1
 /product="deoxynucleoside kinase"
 /protein_id="CAB41881.1"
 /db_xref="GI:4741228"
 /translation="MAEASCARSGTGYAEGTOPFTVLIENISGKTTYLNHEKX
 NDICLTPEVKRWNVNGVNLLETKYDKPKKAMPFOSIVYTLMLQSHTPATPKKI
 MERSTFSAKYCTENMRNGSLBGMNTLEWKFTLESHVQADLIILAKSPYEA
 VERIQKARSESCVPLKYLDELHLEHDLIHLRRPOSKVILVDADLMLNIGTEY
 QNSESIDALISSNQPSPLVSPSKRQVAR"

BASE COUNT 188 a 213 c 214 g 135 t
 ORIGIN

Query Match 98.5%; Score 742; DB 6; Length 750;
 Best Local Similarity 99.3%; Fred. No. 6.7e-18;
 Matches 745; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 atggcggagcagcatcctctgtccggaagggagcaagtaagcggagggagccaccc 60
 Db 1 ATGGCGGAGCAGCATCTCGTCCGGAAGGGAGCAACGTCGCCGAGGGACCCACCC 60
 QY 61 ttacacgtctcatatcgagggcaacatcgagcgggaaagacacgtatttgaacattc 120
 Db 61 TTACACGTCCTCATCGAGGGCAACATCGGAGCGGGAAGACACGTAATTGAACACTTC 120
 QY 121 gagaagtaacaagaacacattgctcgtctgaacgagcccgcgaggaagtgcgaacgtc 180
 Db 121 GAGAAGTACAAAGAAGACATTTGCTGCTGACCAAGCCCGTCGAGAGTGGCGCAACGTC 180

QY 181 aacggggtaaatctgtcgtgagctgtatgtacaagaatcccaagaagtggccatgccctt 240
 Db 181 AAGGGGTAAATCTCTGTGAGCTGTATGTACAAAGATCCCAAGAAGTGGCGATGCCCTTT 240
 QY 241 cagaattatgtacagctgacatgtcgtcagtcgcacacccgcccacaagaagctga 300
 Db 241 CAGAGTTATGTACAGCTGACCATGCTGTCGAGTCGCACACCGGCCCAACCAAGAACTGA 300
 QY 301 aaataaraggaagcgctccatttttaagctcgtcgtcattgtctgtgagaacatgcagca 360
 Db 301 AAATAATAGAGCGCTCCATTTTAAAGCTGCTGCTTCTGTGGAGAACATGCAAGCA 360
 QY 361 aacggctcgtctgagacgaagcgtatcaatacgtctgaggagtggtatacaattcatcgaa 420
 Db 361 AACGCTCGTCTGAGACAGAGGCGCATATCAATATCTGAGAGAGTGTACAGATTTCATCGAA 420
 QY 421 gattccattacacctgcgagcggaacctatcatatctgcacacctgcgcgagagtgagc 480
 Db 421 GATTCATTACAGCTGCAAGCGCGACCTCATATATCTGCCACTTGC CGGAGGTGGCG 480
 QY 481 tacgaacgcatccgagcgagcgctgtctgaggagagcgctgagcgttaagtacct 540
 Db 481 TACGAACGCATCCGGCAGCGGCGCTGTTGAGAGAGAGCTGCTGCGCTTAAGTACCTT 540
 QY 541 caggaagctgtacagttgtgacacgaagactgtgtatcacacagagacgacgacgtctgc 600
 Db 541 CAGGAGCTGCATGAGTGTGACAGAGACTGTTGATFACACAGAGACGACGACGTCTGC 600
 QY 601 aagtcctagtcctcgtatgctgcagatcgtgaaccttgaacatctggacgagtaacagcgc 660
 Db 601 AAGTCTCTAGTCTCTGATGCTGCAGATCTGAACCTGGAATAATGGCACCGATACAGCGCC 660
 QY 661 tcggaagacagcatatcgcagccatctcaagtaacacagacccctgcggttcgtgtg 720
 Db 661 TCGGAGACAGCATATTCGACCCCATCTCAAGTAACCAACAGCCCTCGCGTTCGTG 720
 QY 721 tcgccaagacagcccgagaggttcgcagataa 750
 Db 721 TCGCCAGCAAGCCGACAGAGGTCGCGCAGATAA 750

RESULT 3
 LOCUS AF185268 1246 bp DNA INV 29-FEB-2000
 DEFINITION Drosophila melanogaster deoxyribonucleoside kinase gene, complete
 ACCESSION AF185268
 VERSION AF185268.1 GI:5924307
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1246)
 Munch-Petersen, B., Knecht, W., Lenz, C., Sondergaard, L. and Piskur, J.
 Functional expression of a multisubstrate deoxyribonucleoside
 kinase from Drosophila melanogaster and its C-terminal deletion
 mutants
 J. Biol. Chem. 275 (9), 6673-6679 (2000)
 MEDLINE 20158988
 REFERENCE 2 (bases 1 to 1246)
 Piskur, J.
 AUTHORS Direct Submission
 TITLE Submitted (09-SEP-1999) Institute of Molecular Biology, University
 JOURNAL of Copenhagen, O. Farimagsgade 2A, Copenhagen K DK-1353, Denmark
 FEATURES
 source
 1..1246
 /organism="Drosophila melanogaster"
 /strain="Oregon R"
 /db_xref="taxon:7227"
 /chromosome="3R"


```

KYLSTDDDELTYSGKDNAPQDSAKANHLETSTHNSGIYLLNTEIRSSGGLMSSST
DPSNKSMPSCSGGVRKLCCKGCKFLAAL"
join(37389..37519,37616..38532,38615..38648,38703..>38798)
/gene="Rh2"
/product="CT33926"
/db_xref="FLYBASE:FBan0016740"
/db_xref="FLYBASE:FBgn0003248"
<37389..>38798
/gene="Rh2"
/note="CG16740"
/map="91D3-91D4"
/db_xref="FLYBASE:FBan0016740"
/db_xref="FLYBASE:FBgn0003248"
join(37421..37519,37616..38532,38615..38648,38703..38798)
/gene="Rh2"
/note="Rh2 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0016740"
/db_xref="FLYBASE:FBgn0003248"
/protein_id="AAF55601.1"
/db_xref="GI:7300445"
/translation="MERSHLEPEPDLAHSGPRFOAOSGNGSVLDNVLPMALVNP
YMSRAPDPMASKITGLFTLALIMISCCGGVYVYFGGKSLRFPANLVNLFPS
DFCMASQSPVMIINFEYETVYLPLMCDIYAGGSLFGCVSISMCMIAFDYNYIV
KGINPTMTIKTSIMKILFIMMAVFNWPLIGMSAYVFBGNITAGSIDYMRMNP
RSYLITSLFYVYPLFLICYSWFIIAAVNAHKKAREOKKKNVASSLSSEDCCKS
AEGKLAKVALFTISLMEAMWPLVLCFEGFLKIDGLPLPLTWGATFAATSAVNI
VYGIHPKRYIVLIEKPCMCVGNMDEKPKPDAPASDETTSSEADSKA"
complement(<39713..>40045)
/gene="CG14294"
/product="CT33923"
/db_xref="FLYBASE:FBan0014294"
/db_xref="FLYBASE:FBgn0038656"
/evidence=not_experimental
complement(<39713..>40045)
/gene="CG14294"
/map="91D3-91D4"
/db_xref="FLYBASE:FBan0014294"
/db_xref="FLYBASE:FBgn0038656"
/evidence=not_experimental
complement(39713..40045)
/gene="CG14294"
/note="CG14294 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014294"
/db_xref="FLYBASE:FBgn0038656"
/evidence=not_experimental
/protein_id="AAF55602.1"
/db_xref="GI:7300446"
/translation="MAKAVASKNYDDQHRITHTASRAKNVDAQIHDFYVLENOKV
AAREVDSAVETTKCIEVFSGLEQFKDCIEHVTRLHMERLSLQITTKROASG
ASHLKTIW"
complement(<40400..40729,40945..>40953)
/gene="CG14293"
/product="CT33922"
/db_xref="FLYBASE:FBan0014293"
/db_xref="FLYBASE:FBgn0038657"
/db_xref="FLYBASE:FBgn0038657"
/evidence=not_experimental
complement(<40400..>40953)
/gene="CG14293"
/map="91D3-91D4"
/db_xref="FLYBASE:FBan0014293"
/db_xref="FLYBASE:FBgn0038657"
/evidence=not_experimental
complement(join(40400..40729,40945..40953))
/gene="CG14293"
/note="CG14293 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014293"
/db_xref="FLYBASE:FBgn0038657"
/evidence=not_experimental
/protein_id="AAF55603.1"
/db_xref="GI:7300447"

```

```

/translacion="MEATGOSYGVNVAOOSIDMNEOFKIAOVVAQGLNPSVEARN
DANDQDLTKLALSHCEALFKSTQGVLDHKTKVKAVPFGFTALDRLAAEHMPLYGAT
SGAARIGFEC"
complement(join(42879..43420,43496..43595))
/gene="CG14292"
/product="CT33921"
/db_xref="FLYBASE:FBan0014292"
/db_xref="FLYBASE:FBgn0038658"
complement(<42879..>43595)
/gene="CG14292"
/map="91D3-91D4"
/db_xref="FLYBASE:FBan0014292"
/db_xref="FLYBASE:FBgn0038658"
complement(join(42968..43420,43496..43549))
/gene="CG14292"
/note="CG14292 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014292"
/db_xref="FLYBASE:FBgn0038658"
/protein_id="AAF55604.1"
/db_xref="GI:7300448"

```

Query Match 78.6%, Score 592, DB 5, Length 232380;
Best Local Similarity 84.4%, Pred. No. 2,6e-124; Indels 133; Gaps 2;
Matches 748; Conservative 0; Mismatches 5;

QY	1	atggcgagagcagcctcctgctgcccgaagggagcaagtaagcgcgagggcaccgccc	60
DB	119516	ATGGCGAGGAGCATCTCTGTCCCGAAGGGAGACCAAGTACGCGGAGGACCCACCC	119575
QY	61	ttaccgtccatcagagggcaacatcgagcaggggaagccagatattgaaccattc	120
DB	119576	TTACCGTCCCTCATCGAGGCAACATCGGACCGGAGACCATGATTTGAACCACTTC	119635
QY	121	gaaagaataaagaacagacattgctgctgacccgagccctcgagaagtgcgcaagtc	180
DB	119636	GAAAGATACAGAGACATTTGCTGCTGACCGAGCCCGTGAAGAGTGGCGCAAGTC	119695
QY	181	aacgggataaatctgctgtagctgtagtacaagaatcccaagaagtggccatccctt	240
DB	119696	AACGGGATAATCTGCTGAGTGATGTACAAAGATCCCAAGAAATGGCCATCCCTTT	119755
QY	241	cagaagtatgctgcgtgacatgctgagtcgacacgcccacacaagaagcta	300
DB	119756	CAGAGTATGTACGCTGACCATGCTGACATGCGACACCGCCCAACAAAGAGCTA	119815
QY	301	aaataargagcgtccatctttagcgtc	331
DB	119816	AAATATATGAGCGCTCCATTTTACGCGTGAATTCATTCCTCCGCTT	119875
QY	332	-----gctattgctcgt	344
DB	119876	TGACCAAGAAACGCTAAGCAAACTATCCATTACCATCCGCGCCGAGTATTCCTTCT	119935
QY	345	ggaagaacatcgagcaagagcgtcgtcgagcagagcatgtacaatacgttgaagagtg	404
DB	119936	GGAAGAATCGGACGAACGCGCTCGCTGAGCGAGCAATGTACATCGCTGAGAGAGTG	119995
QY	405	gtacaagtatcgaagagtcattcaactgcaagcgcgagacctcatatattctgcgac	464
DB	119996	GTACAAGTTCATCGAAGAGTCCATTACGCTGCGAGGAGCTCATATATATCTCGGCT	120055
QY	465	ctgcgcgagagtgagcgtgacgaacgcatccggaagggagcgtctgtgaagagagtcgt	524
DB	120056	CTGCCCGAGGTGGCTGACGAACGATCCGCGACCGGCGCTGTTGAGAGAGCTGGCT	120115
QY	525	gocgcttaagtaacctcaagagctgtagtgcacccaagagctggtatgatacaccagag	584
DB	120116	GCGCTTAAGTAACTTCAAGAGCTGCATGATGTGACAGAGACTGTTGATACACAGAG	120175
QY	585	acgacgagcagtcgtga	601
DB	120176	ACGACCGAGCTGCGCAAGGTAGTTCTTACACAAGCTCACATCTACCGAAGCGCTAAT	120235

71.68;	Score 538.8;	DB 60;	Length 157133;
83.28;	Pred. No. 3.1e-112;		

Matches	744:	Conservative	0;	Mismatches	9;	Indels	141;	Gaps	6;
Oy	1	atgacgagagcagatccctgtgcgcgaagaagggagccaagtaocgcagagagccaccg							60
Db	64304	ATGGCGGAGGAGCATCTCTGGCCCGAAAGGAGCACAAGTACCCCGAGGCAACCGCC							64363
Oy	61	ttcacgctccatcgcagggcgaacatccggaaggggaagaacacgfatcttgaaacattc							120
Db	64364	TTACACGTCCTCATTCGAGGGCAACATCGGACGGGAGACCACTGATTTGAAACCACTTC							64423
Oy	121	gagaagtaacaaagacagcattctgctctgtaccagagcccgctcgaagaatgtgcgaagtc							180
Db	64424	GGAGAGTACAAAGAAAGACATTTGGCTCTGCTGACCGAGCCCGCTCGAAGAAGTGGCGCAACGTC							64483
Oy	181	aacggggtaaatctgtctgtgagctgtatgtacaagaatcccaagaagtgggccatcgccctt							240
Db	64484	AACGGGGTAAATCTCTCTGAGACTGATGTACAAAGATCCCAAGAGATGGCCATGCCCTTT							64543
Oy	241	cagagttatgtcaagctgacatctgtgtgagtcgcacacgcgcccacaagaagacta							300
Db	64544	CGAGGTATGTCCACTGACCATGCTGTGAGTCGGACACCGGCCCAACCAAGAAGCTA							64603
Oy	301	aaataaaggagcgctccatltttagcgctc-----							331
Db	64604	AAATATATGAGACGCTCCATTTTATAGCCCTGGTAGTCATTAATCCCGTCCGGTT							64663
Oy	332	-----gctatgtctcgt							344
Db	64664	TGACCCGAAAGCCTAAACGAACATTATACCATTTACCCATCGCGCCGACGATTTCTTCTGT							64723
Oy	345	ggagaacatcy--cgacgaaaogcgctcg--tgagcagaggaatgtaaataacgttgag							400
Db	64724	GGAGACATGANNMGACGAAGAACGGCTCGCTTGGACAGGGCATGTGACAAATACGCTGGAGG							64783
Oy	401	agtgtgtacagttcatcgaagagttcatcaacctgcagcgagacctcaatatactgc							460
Db	64784	AGTGTATCAAGTTTCATCGAAGATCCATTCACGTGCGAAGCGACCTCATATATATCTGC							64843
Oy	461	gcaacctgcggaggtgtgcgtacgaagcatcccgcaagcggtctgttctgaagaaggt							520
Db	64844	GCACCTGGCCGGAGGTGGGTACGAACCGCATCCGACAGCGGGCTGTTCTGAAGGAGACT							64903
Oy	521	gcgtgtgcgcttaagtacatccttcaggaactcgaatgagtttcacagagactgtgtataacc							580
Db	64904	GCGTGGCGCTTAAGTACTCTTCAGGAGCTGATGATAGTTGCACAGAGAGATGTTGATACACC							64963
Oy	581	agagaagcagccagctcgtgca-----							601
Db	64964	AGAAACGACGCGAGTCTGTGACAGATGATTCTTACACACGCTCATCTACGAGCCGCT							65023
Oy	602	-----agttcctagtctcgtatgcacatctgaacctgtgaacatly							643

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 102657)
 Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
 Burenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hinkley, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
 Kearney, L., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,
 Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
 Sequencing of Drosophila melanogaster
 Unpublished

TITLE 2 (bases 1 to 102657)
 Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
 Burenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

TITLE Direct Submission
 JOURNAL Submitted (22-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
 COMMENT Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 11, 2000 this sequence version replaced gi:6599757.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bugreport@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 80 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 890: contig of 890 bp in length
 * 891 970: gap of unknown length
 * 971 1602: contig of 632 bp in length
 * 1603 1682: gap of unknown length
 * 1683 2326: contig of 644 bp in length
 * 2327 2406: gap of unknown length
 * 2407 3212: contig of 806 bp in length
 * 3213 3292: gap of unknown length
 * 3293 4089: contig of 797 bp in length
 * 4090 4169: gap of unknown length
 * 4170 4842: contig of 673 bp in length
 * 4843 4923: gap of unknown length
 * 4923 5607: contig of 685 bp in length
 * 5608 5688: gap of unknown length
 * 5688 6356: contig of 669 bp in length
 * 6357 6437: gap of unknown length
 * 6437 7427: contig of 991 bp in length
 * 7428 7507: gap of unknown length
 * 7508 8306: contig of 798 bp in length
 * 8306 8385: gap of unknown length
 * 8386 9164: contig of 779 bp in length
 * 9165 9244: gap of unknown length
 * 9245 10150: contig of 906 bp in length
 * 10151 10230: gap of unknown length
 * 10231 10982: contig of 751 bp in length
 * 10982 11061: gap of unknown length
 * 11062 11746: contig of 685 bp in length
 * 11747 11826: gap of unknown length
 * 11827 12399: contig of 573 bp in length
 * 12400 12479: gap of unknown length
 * 12480 13371: contig of 892 bp in length

13372 13451: gap of unknown length
 * 13452 14048: contig of 597 bp in length
 * 14049 14128: gap of unknown length
 * 14129 15337: contig of 1209 bp in length
 * 15338 15417: gap of unknown length
 * 15418 15907: contig of 1490 bp in length
 * 15908 16987: gap of unknown length
 * 16988 17749: contig of 762 bp in length
 * 17750 17829: gap of unknown length
 * 17830 18763: contig of 934 bp in length
 * 18764 18844: gap of unknown length
 * 18844 19605: contig of 763 bp in length
 * 19607 19687: gap of unknown length
 * 19687 20783: contig of 1097 bp in length
 * 20784 20863: gap of unknown length
 * 20864 22020: contig of 1157 bp in length
 * 22021 22101: gap of unknown length
 * 22101 23471: contig of 1370 bp in length
 * 23471 23550: gap of unknown length
 * 23551 24806: contig of 1256 bp in length
 * 24807 24886: gap of unknown length
 * 24887 25528: contig of 642 bp in length
 * 25529 25608: gap of unknown length
 * 25609 26310: contig of 702 bp in length
 * 26311 26390: gap of unknown length
 * 26391 27155: contig of 765 bp in length
 * 27156 27235: gap of unknown length
 * 27236 27909: contig of 674 bp in length
 * 27910 27989: gap of unknown length
 * 27990 29585: contig of 1596 bp in length
 * 29586 29665: gap of unknown length
 * 29666 30871: contig of 1206 bp in length
 * 30872 30951: gap of unknown length
 * 30952 31776: contig of 825 bp in length
 * 31777 31856: gap of unknown length
 * 31857 33197: contig of 1341 bp in length
 * 33198 33277: gap of unknown length
 * 33278 34449: contig of 1172 bp in length
 * 34450 34529: gap of unknown length
 * 34530 35270: contig of 741 bp in length
 * 35271 35350: gap of unknown length
 * 35351 36805: contig of 1455 bp in length
 * 36806 36886: gap of unknown length
 * 36887 38097: contig of 1212 bp in length
 * 38098 38177: gap of unknown length
 * 38178 39390: contig of 1213 bp in length
 * 39391 39470: gap of unknown length
 * 39471 41255: contig of 1785 bp in length
 * 41256 41335: gap of unknown length
 * 41336 42858: contig of 1523 bp in length
 * 42859 42938: gap of unknown length
 * 42939 44814: contig of 1876 bp in length
 * 44815 44895: gap of unknown length
 * 44896 46297: contig of 1403 bp in length
 * 46298 46377: gap of unknown length
 * 46378 47921: contig of 1544 bp in length
 * 47922 48001: gap of unknown length
 * 48002 50278: contig of 2277 bp in length
 * 50279 50358: gap of unknown length
 * 50359 53204: contig of 2846 bp in length
 * 53205 53284: gap of unknown length
 * 53285 55654: contig of 2370 bp in length
 * 55655 55734: gap of unknown length
 * 55735 57502: contig of 1768 bp in length
 * 57503 57583: gap of unknown length
 * 57584 59074: contig of 1492 bp in length
 * 59075 59154: gap of unknown length
 * 59155 61499: contig of 2345 bp in length
 * 61500 61579: gap of unknown length
 * 61580 63316: contig of 1737 bp in length
 * 63317 63396: gap of unknown length
 * 63397 65599: contig of 2203 bp in length
 * 65600 65679: gap of unknown length

*	65680	67509: contig of 1830 bp in length
*	67510	67589: gap of unknown length
*	67520	69271: contig of 1682 bp in length
*	69272	69351: gap of unknown length
*	69352	71396: contig of 2045 bp in length
*	71397	71497: gap of unknown length
*	71497	73961: contig of 2485 bp in length
*	73962	74041: gap of unknown length
*	74042	77824: contig of 3783 bp in length
*	77825	77905: gap of unknown length
*	77905	81980: contig of 4076 bp in length
*	81981	82060: gap of unknown length
*	82061	87894: contig of 5834 bp in length
*	87895	87974: gap of unknown length
*	87975	88554: contig of 580 bp in length
*	88555	88634: gap of unknown length
*	88635	89263: contig of 629 bp in length
*	89264	89343: gap of unknown length
*	89344	90143: contig of 800 bp in length
*	90144	90223: gap of unknown length
*	90224	90844: contig of 621 bp in length
*	90845	90924: gap of unknown length
*	90925	91384: contig of 460 bp in length
*	91385	91464: gap of unknown length
*	91465	92059: contig of 595 bp in length
*	92060	92139: gap of unknown length
*	92140	92798: contig of 659 bp in length
*	92799	92878: gap of unknown length
*	92879	93342: contig of 464 bp in length
*	93343	93422: gap of unknown length
*	93423	94181: contig of 759 bp in length
*	94182	94261: gap of unknown length
*	94262	94816: contig of 555 bp in length
*	94817	94896: gap of unknown length
*	94897	95527: contig of 631 bp in length
*	95528	95607: gap of unknown length
*	95608	96340: contig of 733 bp in length
*	96341	96420: gap of unknown length
*	96421	96851: contig of 431 bp in length
*	96852	96931: gap of unknown length
*	96932	97619: contig of 688 bp in length
*	97620	97699: gap of unknown length
*	97700	98268: contig of 569 bp in length
*	98269	98348: gap of unknown length
*	98349	98998: contig of 650 bp in length
*	98999	99078: gap of unknown length

gene	4. . 816				
CDS	/gene="rk2"	4. . 816	/gene="rk2"	/EC_number="2.7.1.21"	/function="phosphorylate pyrimidine nucleosides to their 5', monophosphate"
				/codon_start=1	/product="mitochondrial thymidine kinase 2"
				/protein_id="CAC07190.2"	/db_xref="GI:10798606"
				/translation="MLRSLRSMWARSPPRSGGSGSPSLDGACGLVAPRRAMP	/translation="MLRSLRSMWARSPPRSGGSGSPSLDGACGLVAPRRAMP
				DKDRNDKKRAVYCITGNSIAKCTCLEFESNTDQVYLMKPYLKMVYGNHPLSL	DKDRNDKKRAVYCITGNSIAKCTCLEFESNTDQVYLMKPYLKMVYGNHPLSL
				WYHDSRNGKLQTVOLTMIDQHTRPQSPHYMLMRSTYSARYFENVLYRGKMBE	WYHDSRNGKLQTVOLTMIDQHTRPQSPHYMLMRSTYSARYFENVLYRGKMBE
				VDYALISSEFDMIVARNIDVSDLVILVIRTPPELCTQRLMKRCREEEKYIIPMEYLHAH	VDYALISSEFDMIVARNIDVSDLVILVIRTPPELCTQRLMKRCREEEKYIIPMEYLHAH
				RYEELMLVNGSLFPAAPAVYLIEADHLEKMLEPEONRARIITLPEMKKGP"	RYEELMLVNGSLFPAAPAVYLIEADHLEKMLEPEONRARIITLPEMKKGP"
				2352	
polya_site	568 a	592 c	645 g	580 t	
BASE COUNT					
ORIGIN					

Query Match	16.7%	Score 125.6	DB 94	Length 2385
Best Local Similarity	53.7%	Pred. NO. 1.8e-18		
Matches 307	Conservative	0	Mismatches 259	Indels 6
Oy	73	atcgaaggacaacatcgcgcagcgcggaagaccacgataatttgaaccacatccgcaggaatcaag	132	
Db	181	ATTGAGGGCAATATTGCAAGTGGGAAGACGACATGCTCGTGGATTTCTCTCA--ATACA	237	
Oy	133	aacgacatctgcctctctgtaccagagccctcgaagaatgctgcaacgtcaacgtcaaggtgataat	192	
Db	238	ACACACCTCGAGGAGTAAATGAGAGCTGTGCTCAAGTGGAGAAATGTCATGGCCATTAAC	297	
Oy	193	ctgtctgagctgcgatgtgtaacaagaatcccaagaagtggcgcatgccccttcagagtatgctc	252	
Db	298	CCTCTGAGCCTCATGTCATGATGATGACCCGATGGGGGCTTCACACTGCAGACGTAACTGTG	357	
Oy	253	acgctgaccatgctgcagctgcgaaccgcccacaaccaagaagaactaaataargag	312	
Db	358	CAGCTACACATGCTGTGGACACAGACGCCCTCCACAGATGTACCTGTACGGTTATGGAA	417	
Oy	313	cgctccatttttagcgtctgatactgtcttcgttgagaacatgcagacaagcgtcgtg	372	
Db	418	AGCTAATTTTACAGCGCAAGTACATTGTTGTGAAACCTGTATAGAAGTGGGAAGATG	477	
Oy	373	gaacgagggcatgtaacaatacgcgtggaagatggttacaagtatcagaagatccatcc	432	
Db	478	CCCGAGGTGTGACTACGGCATTTCTGTGATGCTGTGACTGTGATGCTGAGAAACATTGAT	537	
Oy	433	ctgcgaaggcgacatcatatataatctgcgaactgcgcgaagtctgagtcgaagaacatc	492	
Db	538	GTCCTGTTGATCTGATATAGTTTANTCTGCAGAACACATCCGGAATCTGTCACAGATTA	597	
Oy	493	cgcgaacgcggtcgtctctgtagaagagctgcgtgcgcttaagtaaccttcagaagctgat	552	
Db	598	AAGATGAGGTGCCGGGAGAGGAAAGTATTCGATGGAAATACCTCATCTTATTCAC	657	
Oy	553	gaagtgcacaagaactggtttatataccagaagc---gaaccagatgcgcgaagtgctcta	609	
Db	658	CGGCTCTACGAGAGATGGCTGTGTAACGGGAGCCCTCTTCCAGCTCAGACCCCTTTCTG	717	
Oy	610	gttcctgcagtgcgactcgaacctggaanaact	641	
Db	718	GTCATTGAGGCTGACACACACTTGGAGAAAT	749	

RESULT 11

HSMTTK

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

1953 bp

Hom sapiens mRNA for thymidine kinase, partial.

Y10498

Y10498.2

GI:4456607

mt-tk gene; thymidine kinase.

mRNA

PRI

24-FEB-1999

SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 1953)
AUTHORS	Wang, L., Munch-Petersen, B., Herström Sjöberg, A., Hellman, U., Bergman, T., Jönvall, H. and Eriksson, S. Human thymidine kinase 2: molecular cloning and characterisation of the enzyme activity with antiviral and cytostatic nucleoside substrates
JOURNAL	FEBS Lett. 443 (2), 170-174 (1998)
MEDLINE	99142705
REFERENCE	2 (bases 1 to 1953)
AUTHORS	Wang, L.
TITLE	Direct Submission
JOURNAL	Submitted (13-JAN-1997) L. Wang, Swedish Univ. of Agricultural Sciences, Dept. of Veterinary Medical Chemistry, BMC, Box 575, S-751 23 Uppsala, SWEDEN
REMARK	revised by [3]
REFERENCE	3 (bases 1 to 1953)

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	SOURCE
3 (bases 1 to 1953)	Wang, L.	Direct Submission	Submitted (12-FEB-1998)	L. Wang, Swedish Univ. of Agricultural Sciences, Dept. of Veterinary Medical Chemistry, BMC, Box 575, S-751 23 uppsala, SWEDEN	On Mar 21, 1999 this sequence version replaced g1:2522222.	Location/Qualifiers 1..1953

	polYA_signal	BASE COUNT	480 a	467 c	517 g	489 t	ORIGIN
							/organism="Homo sapiens"
							/db_xref="taxon:9606"
							/tissue_type="brain"
mrna							<1..>1953
							/gene="mt-TK"
gene							1..1953
							/gene="mt-TK"
							<1..699
CDS							/gene="mt-TK"
							/codon_start=1
							/product="thymidine kinase"
							/protein_id="CAA71523.2"
							/db_xref="GI:4456608"
							/translation="VORYAMPPEKDEKEKSVICBESNINASGKTCLEFFSNATV
							VLTPKSKMNRVGRHNPGLIMYHDASRMGLTLQYVOLTMDRTRPROVSVMKMR
							IHSRYIFEVNLYRSGKMEVDVYVLESEFMVTLINMVDVLIYVLRTPETCYOR
							KKRREKEVPLEYLEAIHHLHEEMLIKGSLFPMAPAVLIEADHHEEMLELFEBO
							RDRILTPENRKHCP"
							1906..1911
							/gene="mt-TK"
Query Match	15.5%;	Score 116.4;	DB 93;	Length 1953;			
Best Local Similarity	53.4%;	Pred. No. 2.2e-16;					
Matches 268;	Conservative	0;	Mismatches 231;	Indels 3;	Gaps 1		
QY	74	tcgagggcaacatcgcgacgagcggaagacacgattatttgaaccacattcgagaaagtacaaga	133				
DB	65	TCGAGGCGCAATATTGCGACATGGGAGAGACGACATGCTGGAATTCCTCTCCA--ACGGCA	121				
QY	134	acgacattgctcgtcgtgaccgagcccgctcgagaagatggtcgcaacgtgcaacggggttaatc	193				
DB	122	CAGACGTGAGAGTGTTAACGGAGCCTGTGTCCAAATGGAGAAATGTCCTGGCCCAATC	181				
QY	194	tgtctgagctcgatgataaagaatcccaagaagtggcgatggcccttcattcagagttatgtca	253				
DB	182	CTCTGGGCGGTGATGTCACACGATGCTCTCGCTGGGGCTTTACGCTTACACACTTATGTGC	241				
QY	254	cgtctgacatgctgtagtctcgcaacaccgccccaaaccaagaagaagctaaataargagc	313				
DB	242	AGCTCACACAGTCGAGACGACATACCTCGCTCCAGGTGTCATCTGTACGGTTGATGAGAA	301				
QY	314	gtccacatttagcgtctcgctatgcttgcttgtaggaataatgacgacgaacagcgtcgctg	373				

Db 302 GGTGATTACAGCCGACATATTTGTAGAAAACCTGTATAGAACTGGAGATCC 361
 QY 374 agcaggcactgtacaaatacagcggagagtggtacaaatcctgaagatccatcc 433
 Db 362 CAGAACTGACATATGTAGTTCTGTGCGAATGGTTTGACATCTTGAGAACATGAGAC 421
 QY 434 tgcaggcagaccatcatatatactgcgaccctgcggaggtgcgtacgaagcattcc 493
 Db 422 TGTCTGTGATTGATGATTGATTCCTTGGACCAATCCTGAGACTTGTTCACGAGGTTAA 481
 QY 494 ggcagcggcctcgtctcgtcggagagcgtcgtcgccttaagtaacctcagagctcatg 553
 Db 482 AGAAGGATGACGAGGAGAGGAGATCATTCGCTGGAATACCTGGAAGCAATTCAC 541
 QY 554 agtgcaccagcagctgtgat 575
 Db 542 ATCTCATGAGGAGTGCTCAT 563

RESULT 12

AB046005 3060 bp mRNA PRI 18-JUL-2000
 LOCUS AB046005
 DEFINITION Macaca fascicularis brain cDNA, clone:QcCE-13136.
 ACCESSION AB046005.1 GI:9280065
 VERSION f1s (full insert sequence).
 KEYWORDS Macaca fascicularis adult cDNA to mRNA, clone_1lb:macaque brain cDNA library QcCE clone:QcCE-13136.
 SOURCE Macaca fascicularis
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE 1 (sites)
 AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
 TITLE Isolation of full-length cDNA clones from macaque brain cDNA libraries
 JOURNAL Unpublished (2000)
 REFERENCE 2 (bases 1 to 3060)
 AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
 JOURNAL Direct Submission

COMMENT Submitted (14-JUL-2000) to the DDBJ/EMBL/GenBank databases. Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan (E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111(ex. 2120), Fax: 81-3-5285-1181) URL: http://www.nih.go.jp/yoken/genebank/
 LIB Name: macaque brain cDNA library QcCE
 Lab host: TOP10
 Vector: pME18S-FL3 (Acc.No. AB009864)
 R. Site1: DraIII (CACTGTG)
 R. Site2: DraIII (CACCATGTG)
 Description: 1st strand cDNA was primed with an oligo(dT) primer (ATGTGGCTTTTCTTTTCTTTT); double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTCTGCTTAAAGCTGGC]; 3' end primer [CGACCTGACCTGAGCACA]).
 Location/Qualifiers
 1..3060
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /clone="QcCE-13136"
 /clone_1lb="macaque brain cDNA library QcCE"
 /dev_stage="adult"
 472..1269

FEATURES

source
 CDS
 472..1269

/codon_start=1
 /product="unnamed protein product"
 /protein_id="BAB01587.1"
 /db_xref="GI:9280065"
 /translation="MLRLPRGMALALRCFEPGSGSPASGSRVORGAMPDSKE
 REKEKSVICVEENINASGKTCLEFSSNADIEVLTPYSKMNVRGNPLGMYODA
 SRKGLTQIVVLTMLDRHRCQPVSSVRLEMSRISHSARYIFVNLVRSKMEVDVY
 LSEMFVILNNMDSIDLIVYLRTPNEETGCRORLKRRCREKEVIPLEYEALHHLHEE
 WLKGSIFPVAAVLLVLEADHNERMLQLEFQRRDLRLPRNKKLGP"
 BASE COUNT 721 a 758 c 867 g 714 t
 ORIGIN

Query Match 15.1%; Score 113.4; DB 85; Length 3060;
 Best Local Similarity 52.4%; Pred. No. 9.8e-16;
 Matches 299; Conservative 0; Mismatches 266; Indels 6; Gaps 2;

QY 74 tgcaggcagaccatcgcgagcgggagagccagctatttgaaccatccgagaagataaga 133
 Db 635 TCGAGGCGCAATATTGCAAGTGGGAAGACGACATGCCGTGGAATTCCTCCCA--ATGCCA 691
 QY 134 acgacattgcctcgtcgtaccgagccgctcgaagaatgagcagcgtcaagggtaaatc 193
 Db 692 CAGACATTGAGGTATTAGCGAGCCTGTCTCCCAAGTGGAAATGTCCTGCGCCACAATC 751
 QY 194 tgcctgagcctgagtgatacaagaatcccaagaagtggccatcccttcagagtgatgca 253
 Db 752 CTCTGGCCCTGATGTATACCAAGATGCCCTCTCGCTGGGGCCTTACGCTAACACTTACGTGC 811
 QY 254 cgtgcacatcgtcgcagctgcacaccccccacacaaagaagctcaaaaataargagc 313
 Db 812 AGCTCACATGCTGCGACAGCAGCATCTTGTCTAGGTGTCATCTGTACGCTTGATGAGAA 871
 QY 314 gctccatttttagcgtcgtcgtatcgtcgtgagagacatgacgacaaacggtcgtcg 373
 Db 872 GGTGCAATTCACAGCGCAAGATACATTTTGTAGAAAACCTGTATAGAAATGGGAAGATCC 931
 QY 374 agcaggcagcactgtaatacgtcgtgagagtggtgataagttcagaaagatccatccac 433
 Db 932 CAGAACTGACATATGTAGTTCTGTGCGAATGGTTTGACTGTGATCTTGGAGAACATGACG 991
 QY 434 tgcaggcagaccatcatatatactgcgaccctgcggaggtgcgtacgaagcattcc 493
 Db 992 TGTCCATTGATTGATTGATTGTTTCTTGGACCAATCCTGAGACTTGTTCACGAGGTTAA 1051
 QY 494 ggcagcggcctcgtctcgtcggagagcgtcgtcgccttaagtaacctcagagctcatg 553
 Db 1052 AGAAGGATGACGAGGAGAGGAGAGGATCATTCACCTGGAATACCTGGAAGCAATTCAC 1111
 QY 554 agtgcaccagcagctgtgtatataccagagagcga---ccgagtcgtgcaaggtccatg 610
 Db 1112 ATCTCATGAGGAGTGAGTGCATCAAGGACGCTTTTCCCTGCGACCCCTCTTTCGG 1171
 QY 611 tccctgagtcgagctcgtgaacctggaatacat 641
 Db 1172 TGATCGAGGCTGACCAACCATGAGAGAGAT 1202
 RESULT 13
 LOCUS HSU77088 1960 bp mRNA PRI 24-MAR-1997
 DEFINITION Human thymidine kinase 2 (TK2) mRNA, complete cds.
 ACCESSION U77088
 VERSION U77088.1 GI:1905968
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1960)
 AUTHORS Johansson,M. and Karlsson,A.
 TITLE Cloning of the cDNA and chromosome localization of the gene for human thymidine kinase 2


```

Db 1127 CAGAGTGGACTGATCTGTCGGAGATGTTGACTGGATCTTGAGAACATGACG 1186
Qy 424 tgcagcgagccatcatatatactgcagccctgcggaggtgctgacgaagcctcc 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1187 TGTCTGTTGATTGATTAATTACTTGGACCAATCCGACACTGTTGACGAGGTAA 1246
Qy 494 ggcagcgaggtcgtctctgagagagagctgctgccttaagctcagagagctgcatg 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1247 AGAAGAGATGACGAGGAGAGAGAGATCATTCGCTGGATTAATCTGGAGACATTCACC 1306
Qy 554 agtcgacccagagactggttcat 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1307 ATCTTCATGAGAGAGTGCCTCAT 1328

RESULT 15
MMU133749 1040 bp mRNA ROD 25-AUG-1999
LOCUS Mus musculus mRNA for deoxyguanosine kinase 1.
ACCESSION AJ133749
VERSION AJ133749.1 GI:4877286
KEYWORDS deoxyguanosine kinase 1; dgk gene.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1040)
AUTHORS Petrakis,T.G., Kilstaki,E., Wang,L., Eriksson,S. and Tallanidis,I.
TITLE Cloning and characterization of mouse deoxyguanosine kinase.
JOURNAL J. Biol. Chem. 274 (35), 24726-24730 (1999)
MEDLINE 99386949
REFERENCE 2 (bases 1 to 1040)
AUTHORS Petrakis,T.G.
TITLE Direct Submision
JOURNAL Submitted (18-MAR-1999) Petrakis T.G., Dept of Mammalian Genetics,
Institute of Molecular Biology & Biotechnology FORTH, 1527
Vassilika Vouton, Herakleion, 711 10 Crete, GREECE
FEATURES
source
1..1040
/organism="Mus musculus"
/strain="Balb/C"
/db_xref="taxon:10090"
/tissue_type="spleen"
6..122
/gene="dgk1"
gene 6..839
CDS 6..839
/gene="dgk1"
/function="nucleoside metabolism"
/note="mitochondrial dgk homologue"
/codon_start=1
/product="deoxyguanosine kinase 1"
/protein_id="CAB43121.1"
/db_xref="GI:4877287"
/translation="MAARFLRLRRASFRSPLRNALVDAPHANMDGGPRRLICIE
GNIAVGKSTPYKLMKTHPEWQVATEPIAAMONIQAGAKDGTSKRLGNLEMMYOE
PARMSYTFQTLSPFMSRLKVLQLEPIGRLLQAEKSVRFERSVSDRYIFAKNLFENS
LSDIEWHTYQDMHSFLQDEFANRLILHGFILQASPOYCMERLQVORDREERKGEIAY
LQOLHSQHEHWFINKTKLHFEALQHVPLVDVTEDESENAARQELMGVNTFMRN
L"
BASE COUNT 250 a 268 c 264 g 258 t
ORIGIN

```

Query Match 6.5%; Score 49.2; DB 94; Length 1040;
 Best Local Similarity 48.8%; Pred. No. 0.38;
 Matches 166; Conservative 0; Mismatches 168; Indels 6; Gaps 1;

```

Qy 310 ggcgcctcattttagagctgcgtatgctgctgagagacatgcaggaacggctcg 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GAGAGGTCTGTGTACAGTACAGTATATCTTTCAGAAATCTATTGAAAATGGCTCC 485

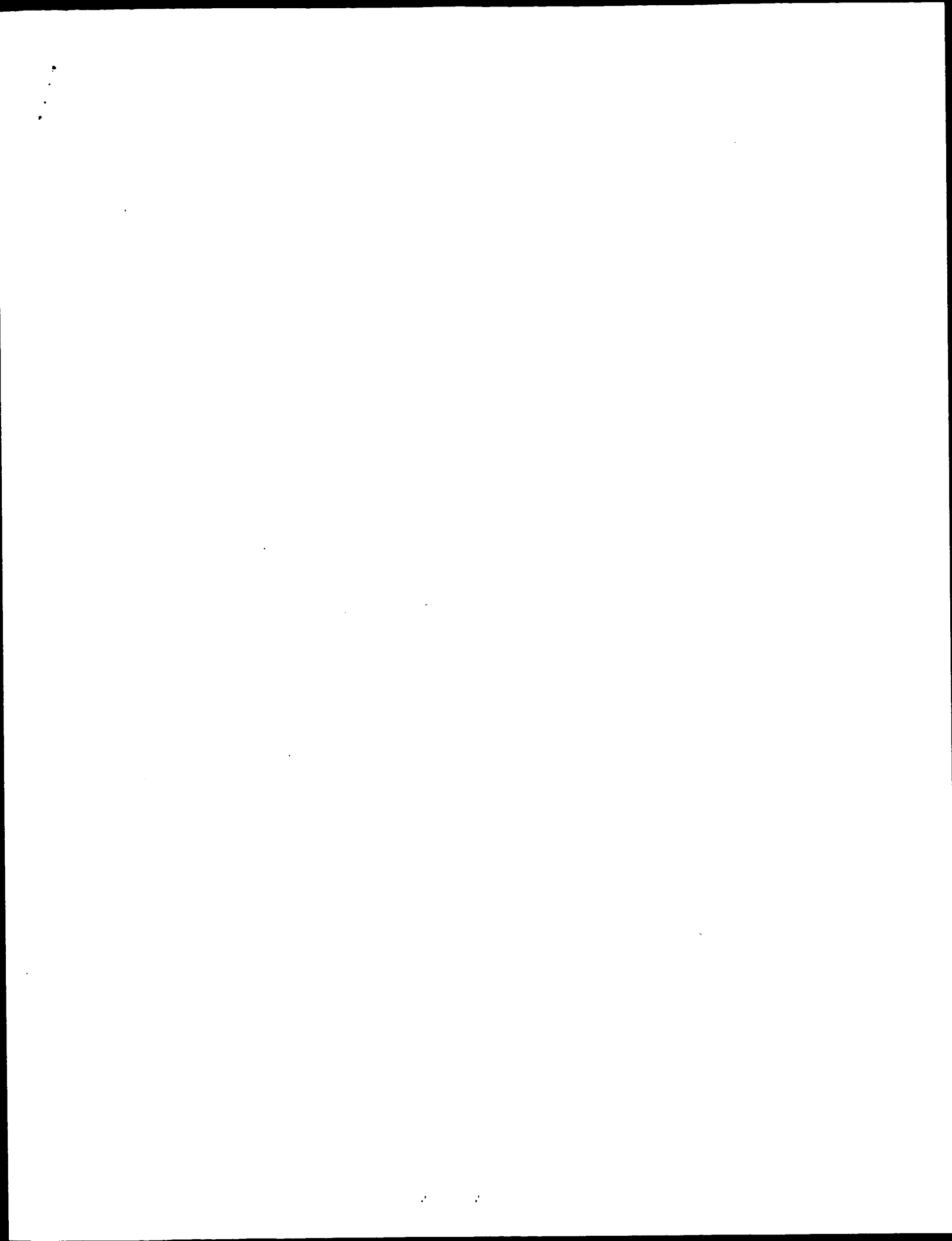
```

```

Qy 370 ctgagcagggcaltgacatacgtctgagagatggtaacatgtaac-----gaagag 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 CTCAGTGCATCTCAGTGGCAGATCATCTATCAGAGACTGGCACTTCTCTCGAGAGATTG 545
Qy 424 tcatcactcagcagcgagcctcatalcatatctgcgacactgcgagagtgagctac 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 GCACACCGGCTTTTGTGTACACGGCTTCATCTATCTCCAGGCTTCAACCCGCTTGTATG 605
Qy 484 gaacgcatccgagcagcgagctcgtctcgtgagagagctgctgcgcttaagfaccttaag 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 GAGAGACTCTACACAGAGGAGATAGAGAAAGAAAGGCAATTGAGCTGGCTTATCTTCAA 665
Qy 544 gagctgcatgagltgcacacagagactggtgtatatacaccagagagacgcaagctgcaag 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 CAGCTACATAGTCCAGCAGCAAGACTGGTTTATTAAACAAGACTAACCAAGCTCCACTTCA 725
Qy 604 gtccatgctcgtgagtgccgatctgaacctgaacctgaagaacatg 643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 GCTCTGACAGACAGTGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765

```

Search completed: October 28, 2001, 06:25:09
 Job time: 6665 sec




```

32 acilysserthrphenalvalasnlleuylsglnleucysglunspirpg 49
143 GCCTGTCACCGCCGCGTGCAGAAAGTGGCGAACGTC..... 180
49 luvalvalprogluprovalalaaargtrpcysasnvalglnserthrln 65
181 .....AACGGGTAAATCTGCT 197
66 aspgluiphegluigluleuthrmetserglunysasnlglyasnvalle 82
198 GGAGCTGATGTACAAAGATCCCAAGAGTGGCGCATGCTTCAGAGTT 247
82 uglmetsmetryglulysprogluargtrpserpherrhphglntrht 99
248 ATGTCAAGCTGACCATGCTGCAGTGCACACCGCCCAACCAACAAGAG 297
99 yralacysleuserarlgilearglalnleualaserleuasnlglyls 115
298 CTAATA.....ATARGAGCGCTCCATTCTTAG 326
116 leuylsaspalaglulysprovalleuuphepnegluargservaltryse 132
327 CGCTCGCTATTGCTCGTGAGAACATGCGACGAAACGCTCGCTGGAGC 376
132 raspargtyrillephealaserasnleutyrgluserglucysmetasnlg 149
377 AGGGCATGTACAAATACGCTGGAGAGTGGACAAATTCATCGAAGAGTCC 426
149 luthrglutrprhlletyrglnasprtrphlsasprmetasnngln 165
427 ATT.....CACTGCAGGGGAGCTCATATATATCTGGACCTCGCC 470
166 phegluylinserleuigluleuaspglyllelletyrglnleuqlnlaatrpr 182
471 GGAGCTGCGTCAGAACGATCCGCGAGGGGCTGTCGAGAGAGAGT 520
182 ogluthrcysleuhsarsgiletyrleuargglylthargasnlglnlgln 199
521 GCGTGCCTTAAGTACCTTCAGAGCTGCATGAGTTGCACAGAGCTGG 570
199 lytleproleuglultryleuiglulysleuhslytrylshsglnsertrp 215
571 TTGATACACAGAGACAGCG.....CAGCTGTGCA 602
216 leuenuhsargthrleuylsyrthrasnphesprtyrleuqlnlgulvalpr 232
603 GGTCTTACTCCGATGCGCGCATGTGAACCTGGAAGAAACATTGGCAGCAGT 652
232 oilleuthrleuaspvalasn.....Gluasprhellysasptryst 246
653 ACCAGCGCTCGAGAGACGATATTTCAGCGCATCTCAAGT 693
246 yrgluysleuvalglulysvalylsglnpneleuhserrthr 259
seq_name: /cgrn_6/ptodata/2/iaa/5B_COMB.pep:US-08-879-561-7
seq_documentation_block:
; Sequence 7, Application US/08879561
; Patent No. 5817482
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Guejler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA

```

```

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,561
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0325 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 260 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1480198
US-08-879-561-7

alignment_scores:
Quality: 333.00 Length: 233
Ratio: 2.297 Gaps: 5
Percent Similarity: 62.232 Percent Identity: 32.618

alignment_block:
US-09-416-579A-1 x US-08-879-561-7 ..
Align seg 1/1 to: US-08-879-561-7 from: 1 to: 260

49 GGCACCCAGCCCTTACCGTCTCATCGAGGCGACATGCGCGGGA 98
111 111 111 111 111 111 111 111 111 111 111 111
18 Glnyarglyproaargargleuserlleglulysnillealvalgluly 34
99 GACCAAGTATTGGAACACACTTCGAGAGTACAAAGACATTGCGTGC 148
111 111 111 111 111 111 111 111 111 111 111 111
34 sserthrphenalylsleuethrlysrthryrproglutprhlisvala 51
149 TGACCGAGCCCGTCAGAGAGTGGCGCAACGTCACAGGGGTA..... 189
111 111 111 111 111 111 111 111 111 111 111 111
51 lathgluprovalalatrhrprglinsnillealalagllytrhln 67
190 .....AATCTGCTGGAGCTGATGTACAA 212
111 111 111 111 111 111 111 111 111 111 111 111
68 lysalacysthrallaglnserleuigllyasnleuenuhsprkethryra 84
213 AGATCCCAAGAGAGTGGCGCATGCCCTTTCAGAGTATGTCACGTCGCA 262
111 111 111 111 111 111 111 111 111 111 111 111
84 gluluproalaaargtrpsertrythrphglntrhpheserphenleusera 101
263 TCGTCGAGTGCACACCGCCCAACCAACAAGAGCTA..... 300
111 111 111 111 111 111 111 111 111 111 111 111
101 rgleuylsvalglnleuqlnupropheproglylslsleuqlnalaarg 117
301 .....AAATAARGAGCGCTCATTTTTCAGCGCTGCTTTCCTT 341
111 111 111 111 111 111 111 111 111 111 111 111
118 lysprovalalaglnleuphegluargservaltryserasprgtyrilleph 134
342 CGTGAGACATGCGACGAAACGCGCTCGCTGAGAGCGCATGTACAAATA 391

```



```

134 ealalyasnlrhegluasnspserleuserrllegluprlr151
132 CGGTGAGAGAGTGTACAACTTATC.....GAAGATCCATTACCG 435
151 leryglinsprtrhlserrrheulrtrpdlrphelaserlrgile 167
436 CAGCGGAGCTCATATATATGCGCACCTGCGCGAGAGTGGCTTACGA 485
168 ThrleuInslrlyhelellyleuInslrlnalaserProglvalCysleu 184
486 ACCGATCGGAGCGGCTGTGTGAGAGAGTGGCTGGCTTAACT 535
184 sargleuylrglnalrglulglulglulglulglulglulglulglul 201
536 ACCTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 585
201 yrluendluInleuInslrlnalserProglvalCysleu 217
586 CGACCG.....CAGTCGTGCAAGTCTTATGCTTGA 617
218 ThrlyleuInslrhegluInslrlnalserProglvalCysleu 234
618 TGCGATCTGAACTGAAACATGCGACGAGACGAGCGGCTGAG 666
234 rylasn.....AspAspserleuInslrlnalserProglvalCysleu 247

seq_name: /cgn2_6/prodata/2/1aa/5B_COMB.pep:US-08-879-561-1

seq_documentation_block:
: Sequence 1, Application US/08879561
: Patent No. 5817482
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Hawkins, Phillip R.
: APPLICANT: Guejler, Karl J.
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/879,561
: FILING DATE: Herewith
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0325 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 261 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

```

: IMMEDIATE SOURCE:
: LIBRARY: FTBROT01
: CLONE: 56821
: US-08-879-561-1

alignment_scores:
  Quality: 287.50      Length: 246
  Ratio: 2.054         Gaps: 7
  Percent Similarity: 56.911  Percent Identity: 29.268

alignment_block:
US-09-416-579a-1 x US-08-879-561-1 ..

Align seg 1/1 to: US-08-879-561-1 from: 1 to: 261

49 GGCACCCACCCCTTCACCCCTTCATCGAGGCAACATCGGACCGGAA 98
18 GATGAGGlyProArgArgleuSerleuInslrlnalserProglvalCysleu 34
99 GACACGATTTGAAACACCTTCGAGAGATGATGATGATGATGATGATGAT 148
34 sserThrPheVallyleuInslrlnalserProglvalCysleu 51
149 TGACGAGCGCGCTGAGAGAGTGGCGGACGTCACGCGGCTA..... 189
51 lArThrGluProValAlaThrTrpGlnAsnllleGlnAlaAlaGlyThrGln 67
190 .....AATCTGCTGAGCTGATGATGATGATGATGATGATGATGATGAT 212
68 LysAlaCysThrAlaGlnSerleuInslrlnalserProglvalCysleu 84
213 AGATCCCAAGAGAGTGGCGGATGCGGCTTCAGAGTATGTCACGTCGACA 262
84 gclurProAlaArgTrpSerlyrThrPheGlnThrPheSerPheleuSer 101
263 TGCTGAGCTGCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
101 rgleuInslrlnalserProglvalCysleuInslrlnalserProglvalCysleu 117
301 .....AAATTAAGAGCGCTTCATTTTAAAGCTGCTGCTGCTGCTGCT 341
118 LysProValGlnllePheGlnArgSerVallyleuSerParglyrleu 134
342 CGTGAGAGACATGCGAGAAAGCGCTGCGGAGAGAGGAGGATGATGATGAT 391
134 ealalyasnlrhegluasnslrlnalserProglvalCysleu 146
392 CGGTGAGAGAGTGTACAACTTATGATGATGATGATGATGATGATGATGAT 441
147 .....lleSerSerlylleSerlylleSerlylleSerlylleSerlylle 155
442 GACCTCATCATTA.....TATCTGCG 461
156 GlylleuPheSerCysGlySerleuProAlaGlySerlnslrlnalser 172
462 CACCTCG.....CCGAGGTGGGCTGAGAGAGGATGATGATGATGATGAT 496
172 aserSerThrSerArgleuProArgPheVallyleuLysleuLysleu 189
497 AGCGGCTGCTTGTGAGAGAGCTGCGGCTTAAAGTACCTGAGAG 546
189 lnaArgAlaArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 205
547 CTCGATGATGTCACGAGAGCTGATGATGATGATGATGATGATGATGATGAT 591
206 leuInslrlnalserProglvalCysleuInslrlnalserProglvalCysleu 222
592 .....CAGTCGTGCAAGTCTTATGCTTGA 628
222 sPheGlnAlaLeuMetAsnllleProValleuValleuAspValasn 237
629 ACCTGAGAAACATTTGCGACCGAGTACCGAGCTGCGAG 666

```

Page 4

cores:		Length:	117
Quality:	164.50	Gaps:	2
Ratio:	2.455		
Similarity:	57.265	Percent Identity:	32.479

Align seg 1/1 to: US-08-860-995-2 from: 1 to: 172

```

49 GGCACCCAGCCCTTCACCGTCATCGAGGCGCAACATCGCCGCGGAA 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18 G1yArgG1yProArgArgLeuSerIleGluG1yAsnIleAlaValG1y 34
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 GACACGATATTGAAACACTTCGAGAGTACAAGAACGACATTGCGTGC 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 sSerThrPheValIyLeuLeuThrIySerIyProG1utPrhIstVal 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 TGACCCAGCCCGTCGAGAGTGGCGGACGTCACCGGGA..... 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 IatPrG1uProValAlaThrPrG1nAsnIleG1nAlaIleG1yThrG1n 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 .....AATCTGCTGAGCTGATGTACAA 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 LysAlaCysThrAlaGlnSerIleuG1yAsnIleuLeuAspMetCysTyr 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 AGATCCCAAGAAAGTGGCCATGCGCTTCAGAGTATATGTACGCTGACCA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 GGIuProAlaArgTyrSerTyrThrPheGlnThrPheSerPheLeuSer 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 TGCTGACGTCGACACCGCCCAACCAACAGAGCTA..... 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 rGleuIyValGlnIleuG1uProPheProG1uIyLeuGlnAlaArg 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 .....AAATTAARGAGCGCTCCATTTTACGCGCTCGCTATTCGCTT 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 LysProValGlnIlePheGlnArgSerValTyrSerAspArgLeuHis 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 C 342
134 e 134

```

seq_name: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:PCT-US95-00532A-2

```

seq_documentation_block:
; Sequence 2, Application PC/TUS9500532A
; GENERAL INFORMATION:
; APPLICANT: WEI, ET AL.
; TITLE OF INVENTION: Human Deoxycytidine Kinase 2
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00532A
; FILING DATE: 13 JAN 95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-246
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1740
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 172 AMINO ACIDS

```

```

; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
PCT-US95-00532A-2

```

```

alignment_scores:
  Quality: 164.50      Length: 117
  Ratio: 2.455        Gaps: 2
  Percent Similarity: 57.265  Percent Identity: 32.479

```

alignment_block:

US-09-416-579a-1 x PCT-US95-00532A-2 ..

Align seg 1/1 to: PCT-US95-00532A-2 from: 1 to: 172

```

49 GGCACCCAGCCCTTCACCGTCATCGAGGCGCAACATCGCCGCGGAA 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18 G1yArgG1yProArgArgLeuSerIleGluG1yAsnIleAlaValG1y 34
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 GACACGATATTGAAACACTTCGAGAGTACAAGAACGACATTGCGTGC 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 sSerThrPheValIyLeuLeuThrIySerIyProG1utPrhIstVal 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 TGACCCAGCCCGTCGAGAGTGGCGGACGTCACCGGGA..... 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 IatPrG1uProValAlaThrPrG1nAsnIleG1nAlaIleG1yThrG1n 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 .....AATCTGCTGAGCTGATGTACAA 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 LysAlaCysThrAlaGlnSerIleuG1yAsnIleuLeuAspMetCysTyr 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 AGATCCCAAGAAAGTGGCCATGCGCTTCAGAGTATATGTACGCTGACCA 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 GGIuProAlaArgTyrSerTyrThrPheGlnThrPheSerPheLeuSer 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
263 TGCTGACGTCGACACCGCCCAACCAACAGAGCTA..... 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 rGleuIyValGlnIleuG1uProPheProG1uIyLeuGlnAlaArg 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 .....AAATTAARGAGCGCTCCATTTTACGCGCTCGCTATTCGCTT 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 LysProValGlnIlePheGlnArgSerValTyrSerAspArgLeuHis 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 C 342
134 e 134

```

seq_name: /cgn2_6/prodata/2/1aa/GB_COMB.pep:US-09-156-836B-2

```

seq_documentation_block:
; Sequence 2, Application US/09156836B
; Patent No. 6242585
; GENERAL INFORMATION:
; APPLICANT: Srivastava, Ranjana
; APPLICANT: Kumar, Deepak
; APPLICANT: Srivastava, Brahm Shanker
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
; FILE REFERENCE: U 011876-4
; CURRENT APPLICATION NUMBER: US/09/156, 836B
; PRIOR APPLICATION NUMBER: 1998-09-18
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (4)

```

[illegible]

WEIR/KEI: UNSURE
LOCATION: (129)

637TTTCCAGTTTCA...TCGGCATCGAG 613

```

663 CGACCGCTGGTACTCG.....TGGCAATGT..... 638
    || ..... ||
112 rocy***seratgatgag***glucyprococysecyseglinh 128
    ..... ||..... ||
637 .....TTTCCAGAGTTCAAA....TCGGATCAG 613
    ..... |||||..... |||||
129 ***Glycylcysargtyrcysargtyrprolysermetgtyrprosemet 145
    ..... |||||..... |||||

```

```

STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1532042
US-08-977-767-3

alignment_scores:
Quality: 99.50 Length: 238
Ratio: 1.19 Gaps: 10
Percent Similarity: 34.874 Percent Identity: 21.008

alignment_block:
US-09-416-579A-1/rev x US-08-977-767-3 ..

Align seg 1/1 to: US-08-977-767-3 from: 1 to: 1345

673 TGCTGCTCTCCGAGCGCTGTACTCGTGCACATGTTTCAGGTTTCAGA 624
|||||: ::::: ||| :||
758 Cyscysalaaglygylglycscycscysalaaathrgly..... 772
623 TCGCATCGAGGACTAGGACCTGGCAGCACTGCCGTCTCTGCTGAT 574
|||||: ||| |:::: :||
773 ..... GlycysCysThrglyglycscysalaatrthrc 784
573 CAACCACTCTGGTGCAACTCATGCACTCTGAAAGTACTTAACGGCA 524
||| |:| ::::: ||||| |||||
784 ysthrcyalaaglyscysthrcysthrrthrglythrcysalaaala 800
523 CGCACTCTCTCCACAAGAGAGCCCGTGGCGGATGCGTTCGTACGCCACC 474
:::::
801 thrglythr..... glyglycysthrglylythrAl 811
473 TCCGGGAGAGTCCGAGATATGATGAGAGTCCGCTCGCAGGTGAATGGA 424
::: |||||
811 aglycsthrlysalaala..... 817
423 CTCTTGATGAAGTACTTACCACCTCTCCAGCGGATGATGACATGCCGTCT 374
::::: ||||| ||| :||
818 ..... AlacyscyscysglyThrThrcyscysthrcysalaagly 830
373 CCAGCAGCGCGTTCTCGCATGTTTCCAGCAAGCAATGACAGCGCTA 324
::::: |||||
831 Alacysglycscysthrcyscysthrrthr..... 840
```

```

323 AAAATGAGCGCTCCYTTATTTTACCTCTTGTTGGTGGCGCGTGTG 274
841 .....AlaAlaLacyscysLgThrglyAlaCysA 850
273 CGACTGACGATGTCAGCGCTGACATACCTGGAAGGCGACCTACT 224
850 LacyscysAlaThrcysAla..... 856
223 TCTTGGGATCTTGTACATCAGCTCCAGAGATTACCCCTTACGTTG 174
857 .....CyscyscysLgThrcysAlaThrcys.....Thrcy 866
173 CGGCACTTCGACGCGCTCGCTGACGA.....GGCAATGTC 136
866 scysThrcAlaThrcAlaAlaAlaThrcGlyAlaThrcGlycyscys 883
135 GTTCTGTACTTCTCGA..... 119
883 hAlaLacysThrcThrcysThrcThrcysAlaAlaGlyAlaCyscysThr 899
118 ...AGTGTCAAAATACGTGCTCTCCGCTGCGGATGTCCTCGANG 72
900 GlyAlaGlycyscysThrglyAlaGlycysThrcyscys,Thr.... 914
71 AGGACGGTGAAGGGCTGGTCCCTGCTACTGTGCTCCCTTGGCGC 22
915 .....Gly 915
21 ACAGATGCTGC 10
916 Thrclycyscys 919

seq_name: /cgn2_6/plodata/2/1aa/5B_COMB.pep:US-08-705-660-46
seq_documentation_block:
; Sequence 46, Application US/08705660
; Patent No. 5858683
; GENERAL INFORMATION:
; APPLICANT: KEESSE, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; NUMBER OF INVENTION: DETECTION OF CERVICAL CANCER
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,660
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: GREENHALGH, DUNCAN A
; REGISTRATION NUMBER: 38,678
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: peptide
US-08-705-660-46

Alignment_scores:
Quality: 99.00 Length: 248
Percent Similarity: 52.823 Gaps: 9
Percent Identity: 24.194

alignment_block:
US-09-416-579A-1 x US-08-705-660-46

Align seq 1/1 to: US-08-705-660-46 from: 1 to: 515

42 CGCGAGAGCGACCCACCTTCACCGCTCTCATTCGAGGAGCAATCGCA 91
71 ArgArgHisGlnThrArgLeuValGlnIleAspAsnGlySglnArgGly 87
92 GCGGAGAGACACGATATTGAAACCACTTCGAGAGTACAGACGACATT 141
87 uphGlnSerArgLeuAlaAspAlaLeuGlnGlnLeuArgAlaGlnHis 103
142 TGCGTCTGACCGAGCCGCTCGAAGAGTGGCGCAACGTCAAGGGTAA 191
104 .....GlnAspGlnValGlnGlnTyrLysGlnLeuGlnLysThrTyr 118
192 TCTGCTGAGAGTGTACAAAGA.....TCCCAAGA 223
119 SerAlaLysLeuAspAsnAlaArgGlnSerAlaGlnArgAsnSerAsn 135
224 AGTGGGCCATGCGCTTTCAGACTTA..... 248
135 uValGlyAlaAlaHisGlnGlnGlnGlnSerArgIleArgIleAspS 152
249 .....TGTACGCTGACCATGCTGACGTCGACACCGCCCAACCA 290
152 erLeuSerAlaGlnLeuSerGlnLeuGlnLysGlnLeuAlaLysGln 168
291 CAAGAGCTAAATAARAGAGCGCTCCATTTT.....AGCG 328
168 uAlaLysLeuArgAspLeuGlnLysSerLeuAlaArgGlnArgAspHis 185
329 CTGCTATTGCTGCTGAGACATGCGAGCAAGCGCTCGCTGAGACAG 378
185 erArgArgLeuLeuAlaGlnLysGlnArgLumetAlaLumetArgAla 201
379 GGCATGTACAATACGCTGAGAGAGTGTACAAGTTCATGAGAGTCAT 428
202 ArgmetGlnGlnGlnLeuAspGlnTyrGlnGlnLeuLeuAspIleLys 218
429 TCACCTGACAGCGCGACCTCATCATAT..... 456
218 uAlaLeuAspmetGlnIleHisAlaTyrArgLysLeuGlnLysGln 235
457 .....CTGCGACCTGCGCGAGGTGGGTACGAGACGATCGGCGAG 498
235 LysIleArgLeuArgLeuSerProSerProHisSerGlnArgSerArg 251
499 CGGCTGCTTCTGAGAGAGC.....TGCCTGCCCT 530
252 ArgAlaSerSerHisSerSerGlnThrglnLysIleValThrcy 268
531 TAAGTACTTTCAGAGCGTCAATGATGTCACACGAGATGTCATGACAC 580
268 sLysArgLysLeuGlnSerThrcySerArgSerSerPheSerGlnHis 285
581 AGAGACGAGCCGACGTGTCGAAAGTCTAGTCTCGATGCCGAT..... 624
285 IaArg.....ThrcyLysArgValAlaValGlnLysArgGln 299
625 .....CTGAACCTGGAACATGCGACCGAGTACACG 657
300 GlyLysPheValArgLeuArgAsnLysSerAsnGlnAspGln 313

```


PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/861,233
 FILING DATE: 31-MAR-92
 ATTORNEY/AGENT INFORMATION:
 NAME: Zeller, James P.
 REGISTRATION NUMBER: 28,491
 REFERENCE/DOCKET NUMBER: 28097/32742
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312/474-6300
 TELEFAX: 312/474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 41:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 355 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-483-533-41

alignment_scores:
 Quality: 98.00 Length: 237
 Ratio: 1.210 Gaps: 10
 Percent Similarity: 34.177 Percent Identity: 23.629

alignment_block:

US-09-416-579a-1/rev x US-08-483-533-41 ..

Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355

```

622 CGGCATCGAGACTAGGACCTTCGACGACTGGCGTCTGCTGTATC 573
    |||||||||
    6 ArgHisArgly.....ProArgArgPr 13
572 AACCATCTCGGTGCACTGACGACCTCTGAGGACTTAAGCGGAC 523
    ||||||| || ||||| |||
    13 OargProProGlyProThrGlyAlaValProThrAlaGlnSerGlnValIT 30
522 GCAGCTCTCTCAGAGAGAGCGCGCT.....GCCGATGCGTTCGTACG 479
    ||||| ||||| ||||| ||||| |||||
    30 hSerThrProAsnSerGlnProAlaValArgSerAlaProAlaAlaAla 46
478 CCACCTCCGCCGAGGTCGCACATATGATGATGCGCGCTGCAGGTA 429
    ||||||| ||||| ||||| ||||| |||||
    47 ProProPro.....ProProAlaGly.. 53
428 ATGCACTCTTCGATGACTGTGACACTCTCCGACGATATTGTACATGC 380
    ||||| ||||| ||||| |||||
    54 .....GlyProProProSerGlySerLeul 62
380 ..... 380
62 euleuArgGlnTrpleuHisValProGlnSerAlaSerAspAspAsp 78
379 .....CTGCTCCAGCGAGCG 363
79 AspAspAspTrpProAspSerProProGlnSerAlaProGlnAlaAr 95
362 TTTCGTCGATGTTCTCCAGCAAGCAATACGACGCTAAATAATGAGCG 313
    ||||| ||||| ||||| ||||| |||||
    95 gProThrAlaAlaAlaProArgProProGlyProHis..... 107
312 CTCCTATTATTAGCTTCTTGTGGTGGGGGGGTGCGACGACGACA 263
107 ..... 107
262 TGCTGACGGTACATACCTGAAGGACATGGCCACCTCTTGATGATCT 213
    ||||| ||||| ||||| |||||
108 .....ArgProAlaTrpAlaArgGlyAlaGlyLe 117
212 TTGTACATCAGCTCCAGAGATTTACCCGCTTGACGCTTGCCGACCTTC 163
    ||||| ||||| ||||| |||||

```

```

117 uThrProProThrProProAlaPro.....SerAlaPhea 130
162 GAGGGCTCGGTACAGCAAGCAATGCTCTGTGACTTTCGAAGGT 113
    ||||| ||||| ||||| ||||| |||||
130 rArgAla...SerProSerAlaCysAlaSerProArgSer..... 142
112 TCAAATAGGTGTCTCCCGCTGCGGATGTTGCCCTGATGAGGACGCTG 63
    ||||| ||||| ||||| ||||| |||||
143 .....ThrTrpArgAla...CysAlaCysAspAlaArgAlaGlyArgG1 156
62 AAGGCTGGGTGCGCTCGCGCTACTGTGTCCTTTCGGGACAGATGC 13
    ||||| ||||| ||||| ||||| |||||
156 yArgArgSerProProArgProProProProProProProProProPro 173
12 TGCCTCGCCCA 2
173 roArgProPro 176

```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US91-06532-3

seq_documentation_block:

Sequence 3. Application PC/TUS9106532

GENERAL INFORMATION:

APPLICANT: Roizman, Bernard

TITLE OF INVENTION: Recombinant Herpes Simplex Viruses

TITLE OF INVENTION: Vaccines and Methods

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

STREET: Two First National Plaza Suite 2100

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/06532

FILING DATE: 19910910

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Gruber, Lewis S.

REGISTRATION NUMBER: 30,060

REFERENCE/DOCKET NUMBER: 27373/8235

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/346-5750

TELEFAX: 312/984-9740

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US91-06532-3

alignment_scores:

Quality: 98.00 Length: 237
 Ratio: 1.210 Gaps: 10
 Percent Similarity: 34.177 Percent Identity: 23.629

alignment_block:

US-09-416-579a-1/rev x PCT-US91-06532-3 ..

Align seg 1/1 to: PCT-US91-06532-3 from: 1 to: 355

```

622 CGGCATCGAGACTAGGACCTTCGACGACTGGCGTCTGCTGTATC 573
    |||||||||

```

```

6  ArgHisArgGly.....ProArgArgpr 13
572 AACAGATCGCTGCATCATGCTAGCTCTGAGGATTAAGCGGAC 523
    ||||| ||| ||||| |||
13  cArgProGlyProThrGlyAlaValProThrAlaGlnSerGlyValr 30
522 GCAGCTCTCTCAGACGAGCCGCT.....GCCGATCCGTTGCGACG 479
    ||||| ||||| ||||| |||||
30  hrSerThrProAsnSerGluProAlaValArgSerAlaProAlaAla 46
478 CCACCTCGCGGAGGTGGCAGATATATGATGAGTCCGCTCGAGGTGA 429
    ||||| ||||| ||||| |||||
47  ProProPro.....ProProAlaGly.. 53
428 ATGAGCTCTCGATGACTTACACTCTCCGACGCTATGTATGATGC. 380
    ||||| ||||| |||||
54  .....GlyProProProSerGlySerLeul 62
380 ..... 380
62  eulLeuArgGlnTrpLeuHisValProGluSerAlaSerAspAspAsp 78
379 .....CCTGCTCCAGCGAGCCG 363
    ||||| ||||| |||||
79  AspAspAspTrpProAspSerProProProGluSerAlaProGluAlaAr 95
362 TTTCGTCGATGTTCCAGCAAGCATAGCGGCTAAATGAGACG 313
    ||||| ||||| ||||| |||||
95  gProThrAlaAlaAlaProArgProProGlyProHis..... 107
312 CTCCCTTATTTTACCTCTGTTGGTGGCGGTGTCGACACTGCACGA 263
107 ..... 107
262 TGGTACGCGTGCATACCTGTGAAGGCGATGCGCCACTTCTTGGGATCT 213
    ||||| ||||| |||||
108 .....ArgProAlaTrpAlaArgGlyAlaGlyLe 117
212 TTGTACATCAGCTCCAGCAGATTACCCGCTGAGCGTTCGACACTTCTC 163
    ||||| ||||| |||||
117  uThrProThrProProArgAlaPro.....SerAlaPheA 130
162 GAGCGGCTCGCTCAGCAGCAATGCTGTTGTAATCTCTCGAAGTGT 113
    ||||| ||||| ||||| |||||
130  rGArgAla...SerProSerAlaCysAlaSerProArgSer..... 142
112 TCATATACGTGCTCTCCCGCTGCCATGTTGCCCTCGATGAGACGGTG 63
    ||||| ||||| ||||| |||||
143  ....ThrTrpArgAla...CysAlaCysAspAlaArgAlaGlyArg 156
62  AAGGCTGGTGCCCTGCGCTACTTGGTCCCTTTCGGGACAGAGATGC 13
    ||||| ||||| |||||
156  yArgArgSerProArgProProArgProProArgProProArgPro 173
12  TGCCCTCCGCA 2
    |||||
173  roArgProPro 176
seq_name: /cgn2_6/ptodata2/1aa/5B_COMB.pep:us-08-660-963-12
seq_documentation_block:
: Sequence 12, Application US/08660963
: Patent No. 5852187
: GENERAL INFORMATION:
: APPLICANT: Thornier, Michael O.
: APPLICANT: Gaylinn, Bruce D.
: APPLICANT: Horikawa, Reiko
: APPLICANT: Lyons Jr., Charles E.
: TITLE OF INVENTION: MOLECULAR CLONING OF THE OVINE PITUITARY
: GROWTH HORMONE RELEASING HORMONE RECEPTOR
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: POPHAM, HAIR, SCHNOBRICH & KAUFMAN, LTD.

```

```

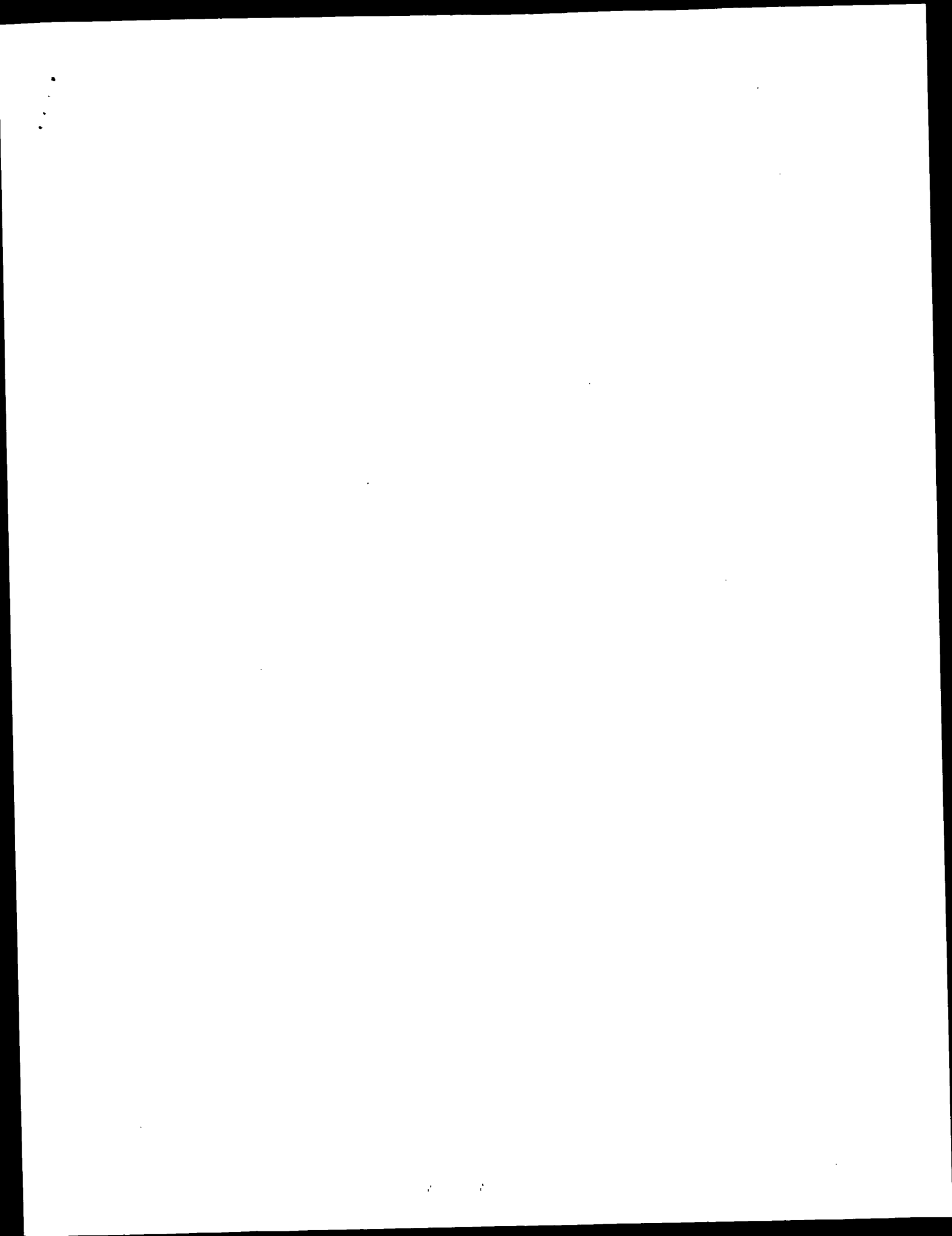
: STREET: Metropolitan Square Building, Suite 800, 1450
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/660,963
: FILING DATE: 12-JUN-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: O'Shaughnessy, Brian P.
: REGISTRATION NUMBER: 32,747
: REFERENCE/DOCKET NUMBER: 18046.036
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-824-8199
: TELEFAX: 202-824-8199
: TELETYPE: 248516
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 498 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-660-963-12

alignment_scores:
    Quality: 96.50      Length: 231
    Ratio: 0.995        Gaps: 10
    Percent Similarity: 41.991    Percent Identity: 24.242

alignment_block:
US-09-416-579a-1/rev x US-08-660-963-12 ..
Align seg 1/1 to: US-08-660-963-12 from: 1 to: 498

604 CCTTGACAGCATCGGCGTCTGCTGATACACAGTCTGTGACAC 555
    ||| ..... ||||| ||||| ||||| |||||
191 ProSerSerSerTrpSerLeuSerGlySerThrAlaProGlyThrTh 207
554 TCATGACAGCTCTGAAGTACTTAAGCGGACGACGCTCTCTCAGAACG 505
    ||| ..... ||||| ||||| |||||
207 r.....SerThrProSerCysS 213
504 AGCCCGCTGCCGA.....TGCGTGTACGCCACCT... 473
    ||||| ..... ||||| ||||| |||||
213 erProProLeuSerSerArgArgGlnLeuSerSerArgThrProProSer 229
472 .....CCGGCGAGTGGCCAGA 456
230 phenrGlyArgThrTrpThrAlaAlaSerProLeuSerCysAlaAr 246
455 TATATGATGAGTCCGCT.....GCAGGTG 430
    ||||| ..... ||||| ||||| |||||
246 gLeu.....LeuProProLeuIleSerArgProProThrSerAlaGlyC 261
429 AATGACATCTTCGATGACTTGTACACCTCTCCAGGATTTGT..... 386
    ||| ..... ||||| ||||| |||||
261 yStrpGlnLeuSerThrProAlaSerProPro.....HisCysProAla 276
385 .....ACATG 381
277 GlnGlyGlySerSerGlyGlyTrpPheSerLeuProGlyGlyPheLeu 293
380 CCCTGCTCAGGAGAGCGGCTTGTGTCGATGTTCTCCAGCAATAGCG 331
    ||||| ..... ||||| ||||| |||||

```


```

|||||
17 ythglnpophetlnValleuilegluGlyAsnilleGlySerIglyst 34
101 CCACGTAATTGAACACCTTCGAGAGTACAGAGACGATTCGCTGCTG 150
|||||
34 hrthrytleuasnhsphneglutytylyAsnaspilleCysleu 50
151 ACCGAGCCGTCGAGAGTGGCGCAACGTCAACGGGGTAATTCGCTGGA 200
|||||
51 ThglnpovaIglulustrpatgAsnValaInglyValaInleu 67
201 GCTGATGTACAAAGATCCCAAGAGTGGCGCATTCGCTGAGATTATG 250
|||||
67 uleuMettylyAspProlyststrPalameProphgInserItyrv 84
251 TCACGCTGACCATGCTGACGTCCGACACCGCCCAACACAAAGAGCTA 300
|||||
84 alThleuthMeleuInserHisThrAlapProthrasInysIstleu 100
301 AAAATARGAGCGCTCCATTTTACGCGCTGCTATTCGCTGAGAA 350
|||||
101 LysileMetGluArgSerIlephSerAlaArgtyrCysPheValgluAs 117
351 CATGGGAGAAACGGCTGCTGAGAGCGCATGTACATATCGCTGGAG 400
|||||
117 nMeLArgAsnnglySerleuInglyInglyMettyrAsnThleuInglu 134
401 AGTGCTACAGTTCTATCGAAGATCCATTCACCTGACGGGAGCGCTCAG 450
|||||
134 lUTrpylyspheIlegluInserIleHisValaInAlaAspIle 150
451 ATPTATCTGCGACCTCGCGGAGGTGGCTACGACGATCGCGGAGCG 500
|||||
151 lIertyleuArgThSerProgluValaIatyrGluArgIleArgInAr 167
501 GGCTGCTGTGAGAGAGCTGCGCTCGCTTAAGACTTACAGAGCTGC 550
|||||
167 galAArgSerGluInserCysValProleuIsttyrleuInleu 184
551 ATGAGTTGCACGAGACTGTTGATACACGAGAGAGCGCAAGTCTGC 600
|||||
184 lsgIleuInhsIgluAspTrpleuIlehsIgluArgdArgProglInserCys 200
601 AAGCTCTAGTCTCGATGCCATCTGAACCTGGAACCATTCGCGACCGA 650
|||||
201 LysValleuValleuAspAlaAspleuAsnleuGluAsnilleGlyThnGl 217
651 GTACACAGCGCTCGAGAGACGATTCGACGCGCATCTCAAGTAACAC 700
|||||
217 urtyeInArgSerIlephAspAlaIleSerSerAsnngIngl 234
701 AGCCCTCGCGGTTCTGCTGTCGCCCAAGAGCGGAGGGTGGCGCAGA 750
|||||
234 lnpProserProValleuValserProserIstArgInArgValaAlaArg 250
seq_name: /SID8/gcgdata/geneseq/AA2000.DAT:AA20726
seq_documentation_block:
ID   AAG20726 standard; protein, 420 AA.
AC   AAG20726;
DT   17-OCT-2000 (first entry)
XX   Arabidopsis thaliana protein fragment SEQ ID NO: 212025.
XX   Protein identification: signal transduction pathway; metabolic pathway;
XX   hybridisation assay; genetic mapping; gene expression control; promoter;
XX   termination sequence.
XX   Arabidopsis thaliana.
OS   Arabidopsis thaliana.
XX   EPI033405-A2.
*PN

```

```

XX   06-SEP-2000.
XX   25-FEB-2000; 2000EP-0301439.
XX   99US-0121825.
XX   99US-0123180.
XX   99US-0123548.
XX   99US-0125788.
XX   99US-0126264.
XX   99US-0126785.
XX   99US-0127462.
XX   99US-0128334.
XX   99US-0128714.
XX   99US-0129845.
XX   99US-0130077.
XX   99US-0130449.
XX   99US-0130510.
XX   99US-0130891.
XX   99US-0131449.
XX   99US-0132048.
XX   99US-0132407.
XX   99US-0132484.
XX   99US-0132485.
XX   99US-0132486.
XX   99US-0132487.
XX   99US-0132863.
XX   99US-0134256.
XX   99US-0134218.
XX   99US-0134219.
XX   99US-0134221.
XX   99US-0134370.
XX   99US-0134378.
XX   99US-0134941.
XX   99US-0135124.
XX   99US-0135353.
XX   99US-0135629.
XX   99US-0136021.
XX   99US-0136392.
XX   99US-0136782.
XX   99US-0137222.
XX   99US-0137502.
XX   99US-0137724.
XX   99US-0138094.
XX   99US-0138540.
XX   99US-0138847.
XX   99US-0139119.
XX   99US-0139452.
XX   99US-0139453.
XX   99US-0139492.
XX   99US-0139454.
XX   99US-0139455.
XX   99US-0139456.
XX   99US-0139457.
XX   99US-0139458.
XX   99US-0139459.
XX   99US-0139460.
XX   99US-0139461.
XX   99US-0139462.
XX   99US-0139463.
XX   99US-0139750.
XX   99US-0139753.
XX   99US-0139817.
XX   99US-0139889.
XX   99US-0140353.
XX   99US-0140354.
XX   99US-0140695.
XX   99US-0140823.
XX   99US-0140991.
XX   99US-0141287.
XX   99US-0141842.
XX   99US-0142154.
XX   99US-0142154.

```

PR	22-SEP-1999;	99US-0155131.
PR	23-SEP-1999;	99US-0155466.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158089.
PR	08-OCT-1999;	99US-0158292.
PR	12-OCT-1999;	99US-0158329.
PR	13-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159283.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159931.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

```

      .               Quality: 315.50      Length: 279
      Percent Similarity: 1.960      Gaps: 8
      Percent Identity: 57.706      Percent Identity: 29.032

Alignment block:
US-09-416-579A-1 x AAG20726 ..

Align seg 1/1 to: AAG20726 from: 1 to: 420

116 ACTTCGAGAGATCAAGAACGACATTTGGCTCGTGACGACGCGCTCAG 161
128 I::: :::: :::: :::: :::: :::: :::: :::: :::: 144
145 LysTrpGlnAspValGlyProAspPheAsnIleLeuAspAlaPheTy 161
166 AAGTGGCGCAGCAGTC.....AAGGGGTAAATCTCTGCGACCTGATGA 205
181 :::: :::: :::: :::: :::: :::: :::: :::: 197
210 CAAGAATCCCAAGAAGTGGGCCATGCCCTTTCAGATGATATGACAGCTGA 255
225 :::: :::: :::: :::: :::: :::: :::: :::: 241
260 CC.....ATGCTCAATCCGACACCGCCGCAACCAACAAACACTTAAA 303
318 hrATgLeMeTgInGlnLysGlnSerAlaSerGlyValLysProLeuArg 194

```



```

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144384.
PR 21-JUL-1999; 99US-0144384.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 28-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 13-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```

```

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

```

Quality: 315.50 Length: 279
Ratio: 1.960 Gaps: 8
Percent Similarity: 57.706 Percent Identity: 29.032

```

alignment_block:

```

US-09-416-579a-1 x AMG20725 ..

```

```

Align seg 1/1 to: AMG20725 from: 1 to: 523

```

```

49 GGCACCCAGCC.....TTACCCGTCATCGA 77
198 GTPGluProAsnLeuAsnValLysLysArgLeuThrPheCysValGl 214
78 GGCACATCGCGCAGCGGAGACCCGATTTG.....AACC 115
214 uGlyAsnLleSerValGlyLysSerThrPheLeuGlnArgIleAlaAsnG 231
116 ACTTCGAGAGTACAGAACGACATTTGCCCTGACCGACCCGCTCGAC 165
231 lUThrValGluLeuGlnAsnLeuValGluIleValProGluProValAsp 247
166 AAGTGGCCGACGTC.....ACGGGTAATCTGCTGAGCTGATGA 209
248 lYsTrpGlnAspValGlyProAspHisPheAsnIleLeuAspLarPhe 264
210 CAAGATCCCAAGATGGGCCATGCCCTTACAGAGTTATGTCAGCTGA 259
264 lSerGluProGlnArgIleValLysLleSerThrPheGlnAsnTyValPheVal 281
260 CC.....ATGCTGACATCCGACACCGCCCAACCAAGAGCTAA 303
281 hArgLeuMetGlnGluLysLleSerIleValLysProLeuArg 297
304 ATAARGAGCGCTCATTTTACGCGCTATGCTGCTGAGACAT 353
298 LeuMetGlnArgSerValPheSerAspArgMetValPheValArgAla 314
354 GCGACGAAGCGCTCGCTGAGCAGGCGATGACATACGCTGAGAGT 403
314 lHisGluAlaLysTrpMetAsnGluMetGluLleSerIleTyrispSer 431
404 GGTACAGTTTCATCGAAGATGCATT...CACCTGACGGCGACCTATC 450

```

```

331 rrpheasprProValSerSerLeuProGlyLeuValProAspGlyPhe 347
||||| :||| |||
501 ATATATTGGCGACCTGCCGAGGTGCCTGACAGCATCCGACGCG 500
||||| :||| |||
348 lIeyrleuArghAlaserProAspRhcyshIsLysArgMetLeuar 364
||||| :||| |||
501 GCGTCGTTTGAGAGAGACTGCGTGGCCTTAAGTACCTCAGAGCTGC 550
||||| :||| |||
364 gLyArGaLaGlucInAspGlyAlSerLeuLysTyrlEuInAspLeu 381
||||| :||| |||
551 ATGAGTTGCACCAAGACTGGTTG..... 573
||||| :||| |||
381 lSGlUyShISglUsertPrLeuPheProPheglUserGlyAasnHisGly 397
||||| :||| |||
574 ..ATACACAGAGACGCGG..... 591
||||| :||| |||
398 ValLeuSerValSerArgProSerLeuHisMetAspAsnSerLeuHisPr 414
||||| :||| |||
591 ..... 591
414 oAsPlleYsasparGyalPheTyrlEuGlUGlYAasnHisMetHisSers 431
592 ....CACTGCTGCAGGCTCTAGCTCTCGATGCCATCTGAACCTGGAA 636
||||| :||| |||
431 erlleGlnLysValProAlaLeuValLeuAspScyluProAsnIlleasp 447
||||| :||| |||
637 .....AACATTGGCACCGAGCTACACAGCGCTCGAGAGCAG 671
||||| :||| |||
448 PheSerArgAsPlleGlnAlaLysThrGlnTyrlAlaArgGlnValAlaGl 464
||||| :||| |||
672 CATATTGCAGCGCCATCTCAAGTAAACCAACAGCCCTCG 708
||||| :||| |||
464 uphePheglUpheValLysLysGlnGluThrSer 476

seq_name: /STD8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG20724

seq_documentation_block:
ID AAG20724 standard; Protein: 580 AA.
XX
XX AAG20724:
AC
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 23023.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter
XX termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-01231825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.

```

PR	30-APR-1999;	99US-0132048
PR	30-MAY-1999;	99US-0132407
PR	04-MAY-1999;	99US-0132444
PR	05-MAY-1999;	99US-0132465
PR	06-MAY-1999;	99US-0132486
PR	07-MAY-1999;	99US-0132487
PR	08-MAY-1999;	99US-0132863
PR	11-MAY-1999;	99US-0134256
PR	14-MAY-1999;	99US-0134281
PR	14-MAY-1999;	99US-0134421
PR	14-MAY-1999;	99US-0134370
PR	14-MAY-1999;	99US-0134370
PR	18-MAY-1999;	99US-0134768
PR	19-MAY-1999;	99US-0134941
PR	20-MAY-1999;	99US-0135124
PR	21-MAY-1999;	99US-0135353
PR	25-MAY-1999;	99US-0135629
PR	25-MAY-1999;	99US-0136021
PR	27-MAY-1999;	99US-0136392
PR	28-MAY-1999;	99US-0136782
PR	01-JUN-1999;	99US-0137222
PR	03-JUN-1999;	99US-0137528
PR	04-JUN-1999;	99US-0137502
PR	07-JUN-1999;	99US-0138094
PR	08-JUN-1999;	99US-0138724
PR	10-JUN-1999;	99US-0138540
PR	10-JUN-1999;	99US-0138847
PR	14-JUN-1999;	99US-0139119
PR	16-JUN-1999;	99US-0139452
PR	16-JUN-1999;	99US-0139453
PR	17-JUN-1999;	99US-0139452
PR	18-JUN-1999;	99US-0139454
PR	18-JUN-1999;	99US-0139456
PR	18-JUN-1999;	99US-0139457
PR	18-JUN-1999;	99US-0139459
PR	18-JUN-1999;	99US-0139460
PR	18-JUN-1999;	99US-0139461
PR	18-JUN-1999;	99US-0139465
PR	18-JUN-1999;	99US-0139730
PR	18-JUN-1999;	99US-0139763
PR	18-JUN-1999;	99US-0139889
PR	21-JUN-1999;	99US-0139817
PR	22-JUN-1999;	99US-0140033
PR	23-JUN-1999;	99US-0140354
PR	23-JUN-1999;	99US-0140695
PR	24-JUN-1999;	99US-0140823
PR	28-JUN-1999;	99US-0140991
PR	30-JUN-1999;	99US-0141287
PR	01-JUL-1999;	99US-0141842
PR	01-JUL-1999;	99US-0142154
PR	02-JUL-1999;	99US-0142055
PR	06-JUL-1999;	99US-0142390
PR	08-JUL-1999;	99US-0142803
PR	09-JUL-1999;	99US-0142920
PR	12-JUL-1999;	99US-0142977
PR	13-JUL-1999;	99US-0143542
PR	14-JUL-1999;	99US-0143624
PR	15-JUL-1999;	99US-0144005
PR	15-JUL-1999;	99US-0144086
PR	16-JUL-1999;	99US-0144085
PR	16-JUL-1999;	99US-0144325
PR	19-JUL-1999;	99US-0144331
PR	19-JUL-1999;	99US-0144332
PR	19-JUL-1999;	99US-0144334
PR	19-JUL-1999;	99US-0144335
PR	20-JUL-1999;	99US-0144352
PR	20-JUL-1999;	99US-0144632
PR	20-JUL-1999;	99US-0144884
PR	21-JUL-1999;	99US-0144814

501 GGCCTGTTCTAGGAGACAGCCTGCCCGCTTAAGAATCCCTTAGAGAGTTC 550
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
421 glysarginalagluglualspgylvalserleuylstrytleuglnasprluh 438

```

551 ATGAGTTGCACCCAGACTGTTG..... 573
||||| |||:|||||
458 ISGLVYSHISGLUSERTRPLeuLeuProPheGLUSERGLYasnHISGLY 454
574 ..ATACACGACGACGACG..... 591
|||||
455 VALLeuSerValSerArgProSerLeuHISMetAspAsnSerLeuHISPr 471
591 ..... 591
471 cAspLeuLysAspArgValPheTYrLeuGLUcLYAsnHISMetHISers 488
592 ....CAGTCGTCGAAGTCCAGTCCTGATGCCGATCGACCTGAA 636
||||| :|||||:|||||:|||||:|||||:
488 eRTLeuGLNValProAlaLeuValLeuAspCYSGluProAsnIleAsp 504
637 .....ACATTGCGACCGAGTACCGACGCTCGAGAGACG 671
||||| :|||||:|||||:|||||:
505 PheSerArgAspIleGLNAlaLysThrGLNTrpAlaArgGLNAlaGL 521
672 CATATTGACGCCATCTCAAGTAAACCAACAGCCCTCG 708
|||||:|||||:|||||:|||||:|||||
521 uPhePheGLuPheValLYSLysGLNGLuThrSer 533
seq_name: /SIDS8/gcdata/geneseq/AA1998.DAT:AAW70493
seq_documentation_block:
ID AAW70493 standard: Protein: 261 AA.
XX
AC AAW70493;
XX
DT 29-DEC-1998 (first entry)
XX
DE Human disease related nucleotide kinase-1 (DRNK-1) protein sequence.
XX
KW Human disease related nucleotide kinase-1; DRNK-1; deoxyguanosine kinase;
KW p21ras; cell proliferation; oncogenesis; cancer; gene therapy;
KW immune disorder; neurological dysfunction.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT peptide 1..25 Mitochondrial localisation signal"
FT modified-site 8 /note= "N-myristoylated"
FT Domain 28..34 /note= "ATP/GTP binding site (P-loop)"
FT modified-site 141 /note= "N-glycosylated"
XX
PN US5817482-A.
XX
PD 06-OCT-1998.
XX
PE 20-JUN-1997; 97US-0879561.
XX
PR 20-JUN-1997; 97US-0879561.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Corley NC, Guegler KJ, Hawkins PR, Hillman JL;
XX
DR WPI; 1998-556388/47.
XX
DR N-PSDB; AAV33481.
XX
PT Nucleic acids encoding deoxyguanosine kinase - useful for
PT recombinant production of the enzyme for treating diseases caused by
PT lack of the enzyme e.g. cancers caused through loss of enzyme
PT function
XX
PS Claim 1; Fig 1A-1C; 53pp; English.

```

```

XX
CC The present sequence represents a human disease related nucleotide
CC kinase-1 (DRNK-1) protein sequence. The DRNK-1 encoding DNA sequence
CC was first identified in Incyte Clone 56821 from the fibroblast cDNA
CC library (FIBRNOR 01). The DRNK-1 DNA sequence is useful for the
CC production of the corresponding recombinant enzyme. The invention
CC provides DRNK enzymes, which are deoxyguanosine kinases, which catalyse
CC the transfer of a terminal phosphate (GMP) to guanosine or guanine in the regulation
CC of cellular levels of GTP and its corresponding nucleoside triphosphate.
CC As GTP levels are known to control the activity of certain oncogenic
CC proteins e.g. p21ras, a protein involved in cell proliferation and
CC oncogenesis, suppression of the enzyme activity causes high ratios of
CC GTP:GDP, promoting oncogenesis. Therefore, diseases (e.g. cancers,
CC immune disorders and neurological dysfunction) related to this lack of
CC gene therapy based strategies. Anti-sense constructs of the DRNK
CC encoding nucleic acids may also be used for inhibition of over-expression
CC of the enzyme.
XX
SO Sequence 261 AA:

```

```

alignment_scores:
Quality: 287.50 Length: 246
Ratio: 2.054 Gaps: 7
Percent Similarity: 56.911 Percent Identity: 29.268

```

```

alignment_block:
US-09-416-579a-1 x AAW70493

```

```

Align seq 1/1 to: AAW70493 from: 1 to: 261

```

```

49 GGCACCCAGCCCTTCCACGTCCTCATCGAGGCGCAACATCGCAGCGGGA 98
||| ||| :|||:|||||:|||||:|||||
18 G|A|A|G|L|P|P|A|A|A|T|P|P|G|L|A|A|S|N|I|L|A|A|V|A|L|G|L|Y 34
99 GACCAGCTATTGACCACTTCGAGAGTACAAGAACGATTCCTCCGCG 148
|||||:|||||:|||||:|||||:|||||
34 sSerThrPheValLysLeuLeuThrLysThrTrpProGLuTrpPheValA 51
149 TGACCGAGCCCGTCGAGAAAGTGCGGCAACGTCACGCGGTA..... 189
|||||:|||||:|||||:|||||:|||||
51 latrGLuProValAlaThrTrpGLNAsnIleGLNAlaLysThrGLN 67
190 .....AATCTGCTGAGCTGATGTAACA 212
|||||:|||||:|||||:|||||:|||||
68 LysAlaCysThrAlaGlnSerLeuGlyAsnLeuLeuAspMetCysTrp 84
213 AGATCCCAAGAGAGGCGCATGCCCTTCAGAGTATGTCACGCTGACCA 262
|||||:|||||:|||||:|||||:|||||
84 gLUProAlaArgTrpSerTrpThrPheGlnThrPheSerPheLeuSerA 101
263 TGCTGCACTGCGACACCGCCCAACAAGAAAGTA..... 300
|||||:|||||:|||||:|||||:|||||
101 rGLeuLysValGLNLeuGLuProPheProGLuLysLeuGLNAlaArg 117
301 .....AAATTAAGGAGCGCTCCATTTTACGCTGCTATTCGTT 341
|||||:|||||:|||||:|||||:|||||
118 LysProValGlnIlePheGlnArgSerValTrpSerAspArgTrpLePhe 134
342 CGTGAGAACATGCGACGAAAGCGCTGCGTGGACGAGGATGTACATA 391
|||||:|||||:|||||:|||||:|||||
134 eAlaLysAsnLeuPheGlnAsnGLysSerLeuSerAsp..... 146
392 CGCTGGAGAGTGTACAAGTTCATGCAAGATCATGCACTGACGAGCG 441
|||||:|||||:|||||:|||||:|||||
147 .....IleSerSerGlyIleSerIleArgThr 155
442 GACCTCATGATA.....TATTCGCG 461
|||||:|||||:|||||:|||||:|||||
156 GlyIleuPheSerCysGLysSerLeuProAlaGLysSerHISTrpMetAl 172

```

```

462 CACCTGG .....CCGAGCTGCCTACGAACCATCCGCC 496
172 aserSerThrSerArgLeuLeuPheProArgPheValIleuLysLysLeuTyrC 189
497 AGCGGGCTGTTCCTGAGAGAGACTGCCTGCCGCTTAAGTACTTTCAAGAG 546
189 IAtgAlaArgGluglInuLysglYrIeGuLeuAlaTyrLeuGlugluIn 205
547 CTGCATGAGTTGCACCAGACTGCTGTATACACCGAGAGACCG..... 591
206 LeuHISglYrInHISgluAlaTrpLeuIIeHISLysThrThrLysLeuH1 222
592 .....CAGTGTGCACAGTGCTCTAAGTCCCTGATGCCGAGCTGA 628
222 sPhelGluAlaLeuMetAsnIleProValIleValIleAspValAsn.... 237
629 ACCTGAACAATGGCACCGACGATACCAGCGCTCGAG 666
238 ....ASPspHeSeTeGlugluJAlaThrLysglIngu 248

seq_name: /SDS8/gcgdata/geneseq/geneseqp/AAL196.DAT:AAW00482
seq_documentation_block:
ID AAW00482 standard; Protein; 172 AA.
AC
XX
AACM00482;
XX
DT 27-SEP-1996 (first entry)
XX
DE Human deoxycytidine kinase 2.
XX
KW Deoxycytidine kinase 2; hdkc2; malignancy; cancer; virus infection; nucleoside analogue; drug activation; therapy; diagnosis.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 28..34
ET /label= ATP-binding_domain
PN W09621724-A1.
XX
PD 18-JUL-1996.
XX
PF 13-JAN-1995; 95WO-USO0532.
XX
PR 13-JAN-1995; 95WO-USO0532.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kirkness EF, Wei Y;
XX
DR WP1: 1996-342279/34.
DR N-PSDB; AAT31065.
XX
XX
DNA encoding human deoxycytidine kinase 2 - useful in treatment of malignancies and viral infections
Claim 1; Page 37; 56pp; English.
XX
XX
Human mature deoxycytidine kinase 2 (hdkc2) (AAW00482) plays a physiological role in the maintenance of normal deoxyribonucleotide pools, and a key role in the phosphorylation of antineoplastic and antiviral nucleoside analogues. It was identified by homology to mouse dck1. It can be expressed in e.g. E. coli, COS or Sf9 cells utilizing vectors incorporating hdkc2 encoding sequences (see also AAT31065). The polypeptide is useful for treating malignancies and viral infections, to screen for antagonists (useful in the treatment of immunodeficiency diseases), and to raise antibodies.
XX
Sequence 172 AA:

```

[illegible]

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 14-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148685.
PR 13-AUG-1999; 99US-0148688.
PR 16-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151338.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 13-SEP-1999; 99US-0154479.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.

```

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 23-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

```

Quality: 150.50 Length: 135
Ratio: 2.120 Gaps: 3
Percent Similarity: 52.593 Percent Identity: 28.889

```

alignment_block:

```

US-09-416-579a-1 x AAG06264

```

```

Align seg 1/1 to: AAG06264 from: 1 to: 202

```

```

433 CTGCGACGGGACCTCATCATATATGCGACCTGCGGAGGTGCGCTA 482
    ||| ||| :|||:|||||:|||||:|||||:|||||:|||||:
21 LeuValProAspIlyPheIleTyrLeuAlaSerProAspIlyCysH1 37
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
483 CGAAGCGATCCGCGACGGGCTGTCGAGAGAGCTGCGCGCTTA 532
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
37 slyArgMetMetLeuArgLysArgAlaGluGlnAspIlyValSerLeuL 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533 AGTACCTTCAGAGCTGCATGATGTCACAGAGCTGGTTG. 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 ystyLeuGlnAspLeuH1sGlnLysH1sGlnSerTrpLeuLeuProPhe 70
574 .....ATACACGAGAGCGACCG..... 591
71 GluSerGlyAsnH1sGlyValLeuSerValSerArgProSerLeuH1sMe 87
591 ..... 591
87 AspAsnSerLeuH1sProAspIlyLeuAspArgValPheTyrLeuGluG 104
592 .....CAGTCGTCGACGAGCTCTAGTCTCGAT 618
104 LysAsnH1sMetH1sSerSerTrpLeuLysValProAlaLeuValLeuAsp 120
619 GCGCATGTGACCTGGA.....ACATTCGACCGACGACGTA 653
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
121 CysGluProAsnIleAspPheSerArgAspIlyGluAlaLysTrpGlnTy 137
654 CGAGCGCTGCGAGCAGCATATTCAGCGCATCTCAAGTAACAACAGC 703
    ||| ||| :|||:|||||:|||||:|||||:|||||:|||||:
137 rAlaArgGlnValAlaGluPheGluPheValLysLysLysGlnGlnT 154
704 CCTCG 708
    |||
154 htrser 155

```

```

seq_name: /SIDS8/gcdata/geneseq/geneseqp/AA1999.DAT.AAY31745
seq_documentation_block:
ID AAY31745 standard; Protein: 430 AA.
XX
AC AAY31745;
XX
DT 22-NOV-1999 (first entry)
XX
DE Mycobacterium tuberculosis specific DNA-encoded polypeptide.
XX
KW Tuberculosis; infection; diagnosis; DNA probe.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 4 /note= "encoded by TGA"
FT Misc-difference 6 /note= "encoded by TGA"
FT Misc-difference 20 /note= "encoded by TGA"
FT Misc-difference 29 /note= "encoded by TGA"
FT Misc-difference 54 /note= "encoded by TGA"
FT Misc-difference 64 /note= "encoded by TGA"
FT Misc-difference 69 /note= "encoded by TGA"
FT Misc-difference 89 /note= "encoded by TGA"
FT Misc-difference 99 /note= "encoded by TGA"
FT Misc-difference 114 /note= "encoded by TGA"
FT Misc-difference 119 /note= "encoded by TGA"
FT Misc-difference 129 /note= "encoded by TGA"
FT Misc-difference 139 /note= "encoded by TGA"
FT Misc-difference 169 /note= "encoded by TGA"
FT Misc-difference 182 /note= "encoded by TGA"
FT Misc-difference 185 /note= "encoded by TAG"
FT Misc-difference 219 /note= "encoded by TGA"
FT Misc-difference 259 /note= "encoded by TGA"
FT Misc-difference 269 /note= "encoded by TGA"
FT Misc-difference 291 /note= "encoded by TGA"
FT Misc-difference 323 /note= "encoded by TGA"
FT Misc-difference 339 /note= "encoded by TGA"
FT Misc-difference 349 /note= "encoded by TGA"
FT Misc-difference 356 /note= "encoded by TGA"
FT Misc-difference 366 /note= "encoded by TGA"
FT Misc-difference 410 /note= "encoded by TGA"
FT Misc-difference 421 /note= "encoded by TGA"
FT Misc-difference 421 /note= "encoded by TGA"
XX
EP945462-A1.
XX

```


composition comprising toluene degrading enzyme useful for biological treatment of organic compounds, especially for degrading toluene or its analogs -

Disclosure: Fig 5; 122pp; English.

CC The present invention relates to toluene degrading enzyme genes and
CC proteins *tutD* (see AAF52629 and AAB59831), *tutI* (AAFF2630 and AAB59832),
CC *tutF* (AAM23631 and AAB52833) and *tutG* (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of p-xylylformate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence for toluene degrading enzyme, *TutD*.
XX
Sequence 1017 AA;

alignment_scores:	
Quality:	137.50
Ratio:	1.058
Percent Similarity:	37.572
	Length:
	Caps:
	15
	Percent Identity:
	23.410

alignment_block;
US-09-416-579A-1 x AAB59813

msg" seg 1/1 to: AAB59813 from: 1 to: 1017

75	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
76	TTTATGATGATGProSerLeuProValThrThAlaSerCysArgProth	91		
77			
78	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
79			
80	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
81			
82	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
83			
84	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
85			
86	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
87			
88	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
89			
90	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
91			
92	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
93			
94	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
95			
96	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
97			
98	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
99			
100	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
101			
102	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
103			
104	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
105			
106	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
107			
108	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
109			
110	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
111			
112	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
113			
114	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
115			
116	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
117			
118	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
119			
120	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
121			
122	TTTTTTTTTTT::::::TTTT	AGGGACCAATACCCGAGG	50
123			
124	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
125			
126	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
127			
128	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
129			
130	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
131			
132	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
133			
134	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
135			
136	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
137			
138	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
139			
140	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
141			
142	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
143			
144	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
145			
146	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
147			
148	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
149			
150	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
151			
152	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
153			
154	TTTTTTTTTTT:::~::~:TTTT	AGGGACCAATACCCGAGG	50
155			
156	TTTTTTTTTTT:::~::~:TTTT		

229 oProSerSerThrArgAsn..... 5 336

... ||:::||
::||:::||
... ||:::||

ATCGAGACATGGACGAACGGCTGCTGGAGCAGC 379

380 GCATTGTAACAATACCAAGCAGCCAGCGAAGTCTTTTC

251ValArgpHeProSerMetalaProLaseRaIqAlaprotH 364

.....G 461

462 CACCTCGCCGGAGGTGGCTACCAACCA

281 IAPR^{Thr}PR^{Arg}Ala^{Arg}Thr^{Pro}Ala^{Thr}Thr^{Pro}Thr^{Pro}Ser^{Ser} 297

.....TCCGGCAGCGGCT 504
||||:||||:.....

505 CGTTCGAGGAGAGCTGCCTGCCCTTAACCTAACCTAACCTA

314 eprOARGrgTThrAlaArgArgCysAlaGlyPheSerSerAlaSerA 331

[illegible]

596 CGTGCAGGTCCTAGTCTCGATGCCGATCTGAACCTTCGAAATGCTTGGGCGG

240 *Algalgasin*..... 351

[illegible]

96 CCAACAGCCCTGCCGGTTCGTGTGTCCGCCAGCAAGC 733

...cyscysalacysargprolaser 379

documentation block:

AB5002C.

4-APR-2001 (first entry)

protein #3 encoded by Tnrd/E gene.

...aromatic
...; iude; yude.

Geobacter metallireducens.

200072650-A2.

- DEC-2000.

TIN: 1000 0000 0000 0000 0000 0000 0000 0000

YOH -) UNIV OHIO.

scungallo PW;

DR WPT; 2001-041080/05.

[illegible]

XX Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs

PS Disclosure: Fig 5; 122pp: English.

XX The present invention relates to toluene degrading enzyme genes and
CC proteins tufH (see AAF23629 and AAB59831), tufT (AAF23630 and AAB59832),
CC tufF (AAF23631 and AAB59833) and tufG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence for toluene degrading enzyme, tufD.

XX Sequence 999 AA:

alignment_scores:

Quality: 132.50 Length: 294
Ratio: 1.123 Gaps: 17
Percent Similarity: 40.136 Percent Identity: 24.830

alignment_block:

US-09-416-579A-1/rev x AAB59817

Align seg 1/1 to: AAB59817 from: 1 to: 999

```

688 AGATGGCGTGAATATGCTGCTCCGAGC..... 659
    ||||| ||| ||| |||||
127 ArgTpsrArg...CysSerSerProSerProArgCysProProse 142
    ||||| ||||| |||||
658 .....GCTGCTACGCTGCCAATGTTCCAGGTTGAG 625
    ||||| ||||| |||||
142 rSerProAlaGlyAlaProGlyAlaThrCysSerArgProPheSera 159
    ||||| ||||| |||||
624 ATGGCATCGAGAGCTAGAGCT.....TGACAG 596
    ||||| ||||| |||||
159 rSerArgAspSerAlaGlyProArgAlaAlaArgPheArgCysArg 175
    ||||| ||||| |||||
595 ACTGGCGTC.....GTCCTGCTGATATCAACGATCCG 561
    ||||| ||||| |||||
176 AspAlaCysGlyAlaArgAlaArgCysProGlyProArgSeraProse 192
    ||||| ||||| |||||
560 TGCACCTATGCACTGCTGCTGAGGACTTAAGCGGCGAGCTCTCTC 511
    ||||| ||||| |||||
192 rLeaArgGlySerArgAspArgSerArgAlaSerArgSera 209
    ||||| ||||| |||||
510 AGAAGAGCCCGCTCCGATGCTGCTGAG.....CCACTCCGCG 467
    ||||| ||||| |||||
209 rGlySerProLeuGlyAlaThrAlaThrSerCysProArgArgArg 225
    ||||| ||||| |||||
466 AGGTGCGCAGATATATGATGAGTCCGCTGAGTGAATGAGACTCTTCG 417
    ||||| ||||| |||||
226 ArgCysSerIle.....GlyAlaSerSeraIle 234
    ||||| ||||| |||||
416 ATGAATGCTGACCACTCTCCA..... 395
    ||||| ||||| |||||
235 .....CysProHisProProValaArgArgSerProValaSerSera 249
    ||||| ||||| |||||
394 .....GCTATTTGACATGCCCTGCTCCAGCAGCCGTTTCGTCGATGT 350
    ||||| ||||| |||||
249 ysaArgAlaHisArgArgCysThrAlaArgGlyArgPheArgGlyPro 265
    ||||| ||||| |||||
349 TCTCCAGAGAACATAGCAGCGC..... 326
    ||||| ||||| |||||
266 ThrSerArgAspThrArgArgArgCysTirArgTirProArgProAr 282
    ||||| ||||| |||||
325 .....TAAATGAGAGCGCTCCTTATTTAGCTTCTT 292
    ||||| ||||| |||||
282 gatGcysArgCysSerArgArgTirPoiArgProleu..... 294

```

```

291 GTTGTGGGGCGGTGTGAGCTCCAGATGCTAGCCTGACATACCTCT 242
    ||| ||| ||| ||| |||
295 ..TTPAlaSerIleCysProArgAlaArgTirArgArgGlySeraSn... 309
    ||| ||| ||| ||| |||
241 GAAAGGCGATGGCCACCTTCTGGGATCTTGTACATCAGCTCCAGACA 192
    ||| ||| ||| ||| |||
310 .....TTPSerSerIleArgSerSeraAlaAla 319
    ||| ||| ||| ||| |||
191 TTTACCCCGTTGACCTTCCGCTTCTGACAGGCTGCTGCA...GCAG 145
    ||| ||| ||| ||| |||
319 r...ProLysArgThrCys.....GlyArgArgValaArgSeraSphrs 333
    ||| ||| ||| ||| |||
144 GCAATGTCTGTTCTTGTACTTCTGAGTGTTCATATAGCTGCTTCC 95
    ||| ||| ||| ||| |||
333 erAlaArgArgSerArgCysProAlaSerSerProIleArgTirArgGly 349
    ||| ||| ||| ||| |||
94 CGCTGCCGA.....TG 84
    ||| ||| ||| ||| |||
350 ArgCysArgArgTirArgArgProleuGlyCysSerProArgAlaThrCys 366
    ||| ||| ||| ||| |||
83 TTGCCCTGATGAGAGCGTGAAGGCTGCTGCTCCCTCGGCTGCTTGT 34
    ||| ||| ||| ||| |||
366 sThrAlaArgCysely.....ArgAspGlyCysSeraAlaPheArgGlyA 381
    ||| ||| ||| ||| |||
33 CCCCTTTCGGGACAGAGATGCTGCTCCGCCA 2
    ||| ||| ||| ||| |||
381 snProleu...HisArgSerLeuArgGlyPro 390
    ||| ||| ||| ||| |||

```

seq_name: /SID8/gcdata/geneseq/geneseq/AA1997.DAT:AA18663

seq_documentation_block:

ID AA18663 standard; Protein: 387 AA.

AC AA18663;

DT 24-JUL-1997 (first entry)

DE Fragmented human NF-H gene +2 frameshift mutant product.

XX Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
XX detection; antibody; probe; cancer; neoplasia; neurodegenerative;
XX Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
XX Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
XX PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
XX cardiovascular; rheumatoid arthritis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 1..387 /note="X corresponds to a stop codon in the
XX accompanying DNA file, AAT69796"

XX MO9712992-A2.

XX 10-APR-1997.

XX 02-OCT-1996; 96WO-1B01106.

XX 11-JAN-1996; 96US-0009832.

XX 02-OCT-1995; 95GB-0020080.

XX (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
XX (UYRO-) UNIV ROTTERDAM ERASMUS.
XX (UYDT-) UNIV STATE UTRECHT.

XX Burdach JPH, Grosveld FG, Van Leeuwen FW;

XX WPI: 1997-226235/20.

XX N-PSDB; AAT69795.

XX Use of mutant genes having frame:shift mutation(s) - for developing
XX prods. for the diagnosis, prevention and treatment of associated


```

466 AGGTGGCAGATATATGATAGCTCCGCTGAGTGAATGACTCTTCG 417
      |||||
819 ArgCysSerIle.....GlyAlaSerSerGly..... 827
416 ATGAACTTGTACCACTCCCTCA..... 395
828 .....CysProHisProProValAlaArgSerProValAsnSerSerL 842
394 .....GCATATGTACATCCCTGCTCCAGCAGCCGTTGCTGCAGAT 350
842 ysarGlaHisArgArgCysThrAlaArgGlyArgPheArgGlyPro 858
349 TCTCCACGAGCAATAGCGAGCGC..... 326
859 ThrSerArgAspThrGlyArgArgGlyCysTrpArgTrpProArgProAr 875
325 .....TAAATGAGCGCTCCCTTATTTTAACTTCTT 292
875 GATGCGArgCysSerArgArgTrpGlyArgProLeu..... 887
291 GTTGTGGGCGGTGCGACCTGACAGCATGGTCAGCATACACTCT 242
888 .....TTPAlaSerGlyCysProArgAlaArgTrpArgArgGlySerAsn... 902
241 GAAGGCGATGGCCCACTCTTGGGATCTTGTACATCAGCTCCACGACA 192
903 .....TTPSerSerGlyArgSerSerAlaAla 912
191 TTTACCCGTTGACGTTGCCCACTTCTGACGCGGCTCGTCA...GCAG 145
912 r...ProLysArgThrCys.....GlyArgArgValArgSerAspThs 926
144 GCAATGCTGCTTGTACTTCTGAGAGTGTCAATATGAGTGTCTTC 95
926 eAlaArgArgSerArgCysProAlaSerSerProIleArgTrpThrGly 942
94 CGCTGCCCA..... 84
83 TTGCCCTGATGAGCAGGTGAAGGCTGGTGGCTCGGCGTACATTGGT 34
939 sThrAlaArgCysGly.....ArgAspGlyCysSerAlaPhePheGly 974
33 CCCCTTGGGCGACAGATGCTCCCTCGGCA 2
974 snProLeu...HisArgSerLeuArgGlyPro 983
seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:AAW18658
seq_documentation_block:
ID AAW18658 standard; Protein: 319 AA.
AC AAW18658;
XX
XX
XX 23-JUN-1997 (first entry)
DE Fragmented human NF-L gene +2 frameshift mutant product.
XX
XX Frameshift mutation product; GAGC motif; somatic mutation; diagnosis;
XX detection; antibody; probe; cancer; neoplasia; neurodegenerative;
XX Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
XX Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
XX PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
XX cardiovascular; rheumatoid arthritis; NF-L; neurofilament-low.
XX
XX Homo sapiens.
XX
XX OS
XX
XX Key Location/Qualifiers
XX FH Misc-difference 1..319
XX FT /note- "X corresponds to a stop codon in the
XX FT accompanying DNA file, AAT69794"

```

```

XX
XX W09712992-A2.
XX
XX 10-APR-1997.
XX
XX 02-OCT-1996; 96MO-IB01106.
XX
XX 11-JAN-1996; 96US-0009832.
XX
XX 02-OCT-1995; 95GB-0020080.
XX
XX (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
XX (UYRO-) UNIV ROTTERDAM ERASMUS.
XX (UYUT-) UNIV STATE UTRECHT.
XX
XX Burbach JPH, Grosveld FG, Van Leeuwen FW.
XX
XX WPI: 1997-226235/20.
XX
XX N-PSDB; AAT69793.
XX
XX Use of mutant genes having frame:shift mutation(s) - for developing
XX prods. for the diagnosis, prevention and treatment of associated
XX diseases, e.g. cancer or neurodegenerative disease
XX
XX Claim 22; Fig 7; 123pp: English.
XX
XX AAW18657 and AAW18658 are +1 and +2 frameshift mutations, respectively,
XX of a sequence comprising fragments of the coding sequence of the
XX human neurofilament subunit NF-L (NF-L) gene corresponding to
XX nucleotides (nt) 1-420 followed by nt 901-1440 of the wild-type NF-L
XX gene. These regions of the gene contain GAGC motifs. Frameshift mutants
XX of the tau, tubulin, apolipoprotein E, microtubule-associated protein
XX 2 (MAP-2), neurofilament subunit L, M and H and amyloid A4 genes are
XX claimed. All these genes share a common GAGC motif (N- A, G, C or T),
XX which is the site of common GA dinucleotide deletion(s) that cause
XX CC neurodegenerative disorders. Antigenic peptides used for the production
XX of antibodies, and small nucleic acid sequences derived from frameshift
XX CC mutants are used in the diagnosis, prevention and treatment of cancer
XX CC disease, Down's syndrome, frontal lobe dementia (Pick's disease),
XX CC Huntington's disease, multiple sclerosis, and other degenerative
XX CC diseases such as cardiovascular disease and rheumatoid arthritis.
XX
XX
XX SQ Sequence 319 AA:

```

```

alignment_scores:
Quality: 121.50 Length: 285
Ratio: 1.038 Gaps: 14
Percent Similarity: 41.053 Percent Identity: 25.614

```

alignment_block:

US-09-416-579a-1 x AAW18658 ..

```

Align seg 1/1 to: AAW18658 from: 1 to: 319
35 CCAAGTACGCGAGGAGGACCCGCTTACCGCTCTCATGAGGCGAC 84
      ||||| ||| |||||
64 ProSerLeuProCysProLeuSerProGlySerHisArgArgGlyAl 80
85 ATCGCAGCGGAGGAGACCAAGTATTTGAACCACTTCGAGAAATCAGAA 134
      : |||||
80 apRoAlaAlaAsnGln**ValProSerAlaThrSerArgThrThrArg 96
135 CGCACTTTCGCTGCTGACCGAGCCGCTGAGAAAGGCGGCAAGCTCAAG 184
      |||
97 .....ProProThrSerGlyAlaThrThrArg 105
185 GGSTAAATCTGCTGAGCTGATGACAAAG..... 214
106 ArgProGlyCysIleSerAlaCysAlaAlaAlaThrAlaProHisAlaG 122
215 .....ATCCCAAGAGTGGCGCATGCC 236

```